SELF-CONSISTENT WASSERSTEIN REPAIR UNDER DECISION-DEPENDENT CONSTRAINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Constraint-enforcing post-processing is usually evaluated at the decision held fixed while the repair is computed. In batch decision systems, however, repaired scores can determine a new threshold; when that threshold also parameterizes the constraints, feasibility becomes endogenous to the repair itself. We formalize this repair–decision feedback and define *self-consistent Wasserstein repair* (SCWR) as a best-response equilibrium between a groupwise Wasserstein projection and the decision induced by the repaired scores. For grouped one-dimensional scores with polyhedral constraints, the fixed-decision repair is a weighted isotonic quadratic program. We derive an exact identity that separates frozen-problem optimality from deployment violation, prove equilibrium existence without contraction, and give a computable global certificate for uniqueness and robustness. Locally, equilibrium sensitivity is governed by the resolvent $(I - J)^{-1}$ of the repair–decision Jacobian; for a scalar decision the feedback amplification is exactly $1/|1 - \chi|$ after factoring out the direct score response. A critical-cone formula covers active-set changes. Under calibration-selected policies with nontrivial feedback, one-shot repair becomes infeasible after re-decision in 94–100% of runs in the main benchmark table, while converged SCWR runs reduce the smooth-decision residual below 10^{-9} with $|\Delta\mathrm{AUC}| < 10^{-3}$. Across 169 stable perturbation trials, the predicted scalar amplification matches the empirical ratio with 0.59% mean relative error.

1 INTRODUCTION

A post-processing rule can be mathematically exact when it is computed and already invalid when it is deployed. Consider a batch system that accepts a fixed fraction of cases. A repair changes the scores to satisfy a risk, capacity, or group constraint at the current threshold. Deployment then recomputes the threshold from the repaired scores. If that same threshold enters the constraint, the repair is evaluated against a different feasible set. The failure is therefore not an optimization error or a distribution shift: the repaired output changes the rule that defines feasibility.

This paper studies that *endogenous-feasibility* loop. The central question is: *what should count as a valid repair when the repair changes the decision that defines its own constraints?* The mechanism appears whenever decisions such as rejection thresholds, top-k cutoffs, or triage boundaries are recomputed from repaired scores. Figure 1 contrasts a frozen one-shot repair with the self-consistent object studied here.

Let s denote the original scores, d the downstream decision, and $\mathcal{K}(d)$ the decision-indexed feasible set. The minimum-distortion repair at a frozen decision is

$$P(s, d) := \arg \min_{q \in \mathcal{K}(d)} \frac{1}{2} \|q - s\|_W^2. \tag{1}$$

A deployment-consistent repair is a pair

$$q^\star = P(s, d^\star), \qquad d^\star = \Theta(q^\star). \tag{2}$$

Thus the repaired scores are feasible for the decision they themselves induce. Equation (2) is a projection fixed point, equivalently a quasi-variational inequality (QVI), because the feasible set depends on the solution.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

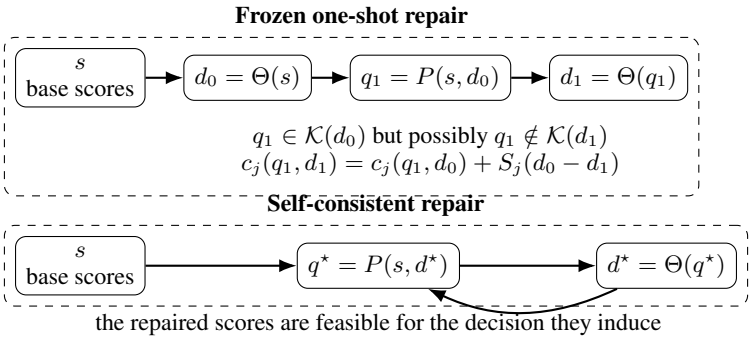

Figure 1: The deployment-consistency gap. A one-shot projection can be exact for the frozen decision d_0 yet violate a moving constraint after re-decision. The equality in the upper panel is Theorem 1; c_j is the signed moving-row residual defined in Section 4, and the lower loop is exactly (2).

The solution concept is operational. If a modular system alternates minimum-distortion repair at the current decision with re-decision, every stationary state satisfies (2). This differs from directly minimizing distortion over the self-referential set $\{q : q \in \mathcal{K}(\Theta(q))\}$: the latter redesigns repair and decision as one global nonconvex program, whereas SCWR characterizes the equilibrium of the repeated repair–decision loop. Section 9 compares the two.

Local behavior is determined by

$$J := D\Theta(q^\star) D_d P(s, d^\star), \qquad \Gamma_{\mathrm{fb}} := \|(I - J)^{-1}\|_2. \tag{3}$$

The spectral radius $\rho(J)$ governs local repeated-repair dynamics, while Γ_{fb} governs equilibrium amplification. When $p = 1$, J is scalar; we denote it by $\chi := J$. The corresponding feedback factor is $1/|1 - \chi|$. Solver choice can change the cost of locating an equilibrium but not this intrinsic branch sensitivity.

Our contributions are threefold. First, we isolate endogenous feasibility and derive an exact one-shot deployment-residual identity, separating frozen optimization error from re-decision error. Second, we develop the corresponding equilibrium theory: existence without contraction, a global uniqueness/robustness certificate, exact regular-stratum sensitivity, and a critical-cone directional formula at active-set changes. Third, we exploit the one-dimensional transport structure to obtain a CPU algorithm and test the theory under calibration-frozen policies, cross-state transfer, smooth/hard boundary audits, a direct joint baseline, and perturbation/gradient checks.

2 Related Work and Positioning

One-dimensional Wasserstein projection reduces to convex optimization in quantile coordinates (Villani, 2009; Peyré and Cuturi, 2019). Existing work covers fixed quantile constraints (Il Idrissi et al., 2024), shape-constrained Wasserstein sets (Matsuda and Wong, 2026), tangential projections between measure sets (Gunsilius et al., 2024), and OT projection for integrating coupled and marginal information (Kim et al., 2026). These formulations do not contain the repair–decision closure in Figure 1.

Performative prediction studies decision-induced changes in the data distribution (Perdomo et al., 2020; Mendler-Dünner et al., 2020). Decision-dependent constrained optimization allows both distributions and linear constraints to respond to the decision and analyzes equilibrium and optimal solutions (Wang et al., 2025); Wasserstein-robust performative prediction adds model-dependent ambiguity sets (Wang et al., 2026). Recent work also studies sensitivity and convergence of performatively stable equilibria under general distribution maps (Zhong et al., 2026). The distinction here is structural: the base sample is fixed, an inner minimum-transport score repair changes the prediction distribution, and the induced decision moves the feasible set of the next repair. We therefore study conditioning of a nested projection, not sensitivity of a decision-induced data-distribution map.

The fixed-point, implicit-differentiation, and variational-sensitivity tools we use are standard (Amos and Kolter, 2017; Bai et al., 2019; Blondel et al., 2022; Rockafellar and Wets, 1998; Shapiro, 2016).

Table 1: Positioning of the feedback object. A check mark denotes a defining component of the formulation, not an empirical advantage.

Setting	Data law responds	Moving constraints	Metric repair
Fixed score post-processing	–	–	optional
Performative prediction	✓	–	–
Decision-dependent constrained opt.	✓	✓	–
SCWR (this paper)	–	✓	✓

We do not claim novelty for these tools or for one-dimensional quantile transport. The problem-specific contributions are the exact deployment-residual identity and the repair-specific conditioning created by the nested projection. Wasserstein distance supplies a rank-preserving distortion and a structured solver; it is not presented as new transport geometry.

3 Problem Setup

3.1 Groupwise Wasserstein distortion

There are G groups. Within group g, sort the base scores as $s_{g1} \le \cdots \le s_{gn_g}$. Each atom has within-group probability weight $\omega_{gi} > 0$ with $\sum_i \omega_{gi} = 1$, and group mass $\pi_g > 0$ with $\sum_g \pi_g = 1$. Let $w_{gi} = \pi_g \omega_{gi}$, concatenate the groups into $s \in \mathbb{R}^N$, and set $W = \mathrm{diag}(w_1, \dots, w_N)$. We use $\langle x, y \rangle_W = x^\top W y$ and $\|x\|_W^2 = x^\top W x$.

For a repaired vector q_g with the same within-group ordering, define $\mu_g = \sum_i \omega_{gi} \delta_{s_{gi}}$ and $\nu_g(q_g) = \sum_i \omega_{gi} \delta_{q_{gi}}$, and

$$W_{2,\mathrm{grp}}^2(s, q) := \sum_{g=1}^{G} \pi_g W_2^2(\mu_g, \nu_g(q_g)). \tag{4}$$

This is a product of groupwise Wasserstein costs; it does not allow transport across groups.

Proposition 1 (Empirical transport identity)**.** *If each q_g is nondecreasing in the order of s_g, then*

$$W_{2,\mathrm{grp}}^2(s, q) = \|q - s\|_W^2. \tag{5}$$

The rank-preserving coupling $s_{gi} \mapsto q_{gi}$ is optimal within each group.

The result follows from the one-dimensional quantile representation; Appendix A gives the proof. Let $\mathcal{M}$ be the product of groupwise monotone cones and $\mathcal{Q} = \mathcal{M} \cap [0, 1]^N$ for probabilistic scores.

3.2 Decision-dependent feasible sets

Let $d \in \mathcal{D} \subset \mathbb{R}^p$ be the downstream decision. We consider fixed polyhedral shape constraints and finite-rank moving constraints written as

$$\mathcal{K}(d) = \{q \in \mathbb{R}^N : Cq \le r + Sd\}. \tag{6}$$

Rows of C include monotonicity, boxes, and policy inequalities; rows of S are zero for fixed constraints. The model includes averages over quantile blocks, trimmed/tail means, and group-specific linear score summaries. Whenever $\mathcal{K}(d)$ is nonempty, strong convexity makes the best response $P(s, d)$ in (1) unique.

The downstream map $\Theta : \mathbb{R}^N \to \mathcal{D}$ can be any continuous policy map covered by the assumptions below. The experiments use a smooth capacity threshold: for coverage $\alpha \in (0, 1)$ and temperature $\tau > 0$, $d = \Theta_\tau(q)$ is the unique root of

$$\sum_{i=1}^{N} w_i \sigma\left(\frac{q_i - d}{\tau}\right) = \alpha, \qquad \sigma(u) = \frac{1}{1 + e^{-u}}. \tag{7}$$

The feasibility experiments use this same smooth decision. Hard top-k quantities are audited separately; no hard-threshold differentiability claim is made.

Definition 1 (Self-consistent Wasserstein repair)**.** *A pair $(q^\star, d^\star)$ is an SCWR equilibrium if (2) holds. Equivalently, with $T_s(d) := \Theta(P(s,d))$, $d^\star = T_s(d^\star)$.*

For a fixed equilibrium decision, $q^\star$ minimizes Wasserstein distortion over $\mathcal{K}(d^\star)$, but it need not solve the global problem over $\{q : q \in \mathcal{K}(\Theta(q))\}$. Metric-projection optimality gives the QVI form

$$q^\star \in \mathcal{K}(\Theta(q^\star)), \qquad \langle q^\star - s, z - q^\star \rangle_W \geq 0 \quad \forall z \in \mathcal{K}(\Theta(q^\star)). \tag{8}$$

4 Why One-Shot Repair Fails

A one-shot procedure freezes $d_0 = \Theta(s)$, computes $q_1 = P(s, d_0)$, and stops. Deployment uses $d_1 = \Theta(q_1)$. For row j of (6), define $c_j(q,d) = C_j q - r_j - S_j d$, so $c_j(q,d) \leq 0$ means feasibility.

Theorem 1 (Exact self-invalidation)**.** *For every moving constraint j,*

$$c_j(q_1, d_1) = c_j(q_1, d_0) + S_j(d_0 - d_1). \tag{9}$$

If the row is active at the frozen repair, $c_j(q_1, d_0) = 0$, then

$$[c_j(q_1, d_1)]_+ = [S_j(d_0 - d_1)]_+. \tag{10}$$

The identity is algebraically simple but operationally decisive: solving the frozen projection more accurately cannot remove a residual caused by moving the feasible set. Equation (10) predicts both the sign and magnitude of that deployment error from the decision shift alone; Section 9 tests this prediction directly.

5 Existence and Global Well-Posedness

Assumption 1 (Continuous bounded feedback)**.** *$\mathcal{D}$ is nonempty, compact, and convex. For every $d \in \mathcal{D}$, $\mathcal{K}(d)$ is nonempty, closed, and convex. The set map $d \mapsto \mathcal{K}(d)$ is inner and outer semicontinuous, the best responses $\{P(s,d) : d \in \mathcal{D}\}$ are bounded, and Θ is continuous with $\Theta(P(s,\mathcal{D})) \subseteq \mathcal{D}$.*

For probabilistic scores, the box $[0,1]^N$ makes boundedness automatic.

Theorem 2 (Existence and QVI characterization)**.** *Under Assumption 1, at least one SCWR equilibrium exists. Every equilibrium satisfies (8); conversely, every solution of (8) together with $d^\star = \Theta(q^\star)$ is an SCWR equilibrium.*

Existence does not imply uniqueness. Outside the contraction regime, the implementation follows a prespecified outer solver and records numerical nonconvergence rather than interpreting it as nonexistence.

Suppose

$$\|P(s,d) - P(s,d')\|_W \leq L_P \|d - d'\|_2, \qquad \|\Theta(q) - \Theta(q')\|_2 \leq L_\theta \|q - q'\|_W, \tag{11}$$

and let $\kappa = L_P L_\theta$. In the perturbation bounds below, W, C, r, S, and Θ are fixed. The vectors s and $\tilde{s}$ are represented on the same groupwise quantile grid: group membership, quantile slots, and atom weights are fixed, although sample identities may cross ranks before re-sorting into those slots. Perturbations that change group composition or empirical weights are outside the statement. In the polyhedral model, L_P is computable. Let $\mathcal{B}$ be the full-row-rank active bases realized on nonempty affine cells of the parametric solution map and define

$$L_P^{\mathrm{poly}} = \max_{I \in \mathcal{B}} \left\| W^{-1/2} C_I^\top (C_I W^{-1} C_I^\top)^{-1} S_I \right\|_2. \tag{12}$$

Proposition 2 (Polyhedral best-response certificate)**.** *For the strongly convex polyhedral projection (1), $d \mapsto P(s,d)$ is continuous piecewise affine. Consequently it is Lipschitz on $\mathcal{D}$, with $L_P \leq L_P^{\mathrm{poly}}$.*

Theorem 3 (Global uniqueness, robustness, and the price of closing the loop)**.** *Assume (11) and $\kappa < 1$. Then the equilibrium is unique and Picard iteration $d_{t+1} = T_s(d_t)$ satisfies*

$$\|d_t - d^\star\|_2 \leq \kappa^t \|d_0 - d^\star\|_2. \tag{13}$$

If s changes to $\tilde{s}$ while W, C, r, S, and Θ remain fixed, then

$$\|\tilde{d}^\star - d^\star\|_2 \le \frac{L_\theta}{1-\kappa}\|\tilde{s} - s\|_W, \qquad \|\tilde{q}^\star - q^\star\|_W \le \frac{1}{1-\kappa}\|\tilde{s} - s\|_W. \tag{14}$$

For any frozen d_0, $q_1 = P(s, d_0)$, and $d_1 = \Theta(q_1)$,

$$\|d^\star - d_0\|_2 \le \frac{\|d_1 - d_0\|_2}{1-\kappa}, \qquad \|q^\star - q_1\|_W \le \frac{L_P}{1-\kappa}\|d_1 - d_0\|_2. \tag{15}$$

The global certificate is deliberately conservative; the next section identifies the exact local sensitivity of a particular equilibrium branch.

6 Intrinsic Feedback Sensitivity

6.1 Regular active-set calculus

Let $(q^\star, d^\star)$ be an equilibrium. We call a neighborhood a *regular stratum* if the solution remains on the same affine face: the active normal span is constant, inactive rows remain strictly slack, and a full-row-rank subset I represents that face. Active rows outside I are locally redundant on the stratum, so their equalities follow from the basis equalities and the affine compatibility relations defining the face. With

$$C_I q = r_I + S_I d, \qquad M_I = C_I W^{-1} C_I^\top, \tag{16}$$

we have the following derivatives.

Proposition 3 (Regular best-response derivatives)**.** *On a regular stratum,*

$$P(s, d) = s - W^{-1} C_I^\top M_I^{-1}(C_I s - r_I - S_I d), \tag{17}$$

$$A_s := D_s P = I - W^{-1} C_I^\top M_I^{-1} C_I, \qquad A_d := D_d P = W^{-1} C_I^\top M_I^{-1} S_I. \tag{18}$$

Let $H = D\Theta(q^\star)$, $B = HA_s$, and $J = HA_d$. Define $\widetilde{B} = BW^{-1/2}$ so that score perturbations are measured in the W-norm.

Theorem 4 (Intrinsic feedback sensitivity)**.** *Suppose $(q^\star, d^\star)$ lies on a regular stratum and $I - J$ is nonsingular. Then the local equilibrium branch is differentiable and*

$$D_s d^\star = (I - J)^{-1} B, \qquad D_s q^\star = A_s + A_d(I - J)^{-1} B. \tag{19}$$

Hence the exact local decision sensitivity (the gain from W-norm score perturbations to Euclidean decision perturbations) is

$$\mathrm{Sens}_d := \|D_s d^\star\|_{W \to 2} = \|(I - J)^{-1}\widetilde{B}\|_2. \tag{20}$$

If $\widetilde{B}$ has full row rank,

$$\sigma_{\min}(\widetilde{B})\Gamma_{\mathrm{fb}} \le \mathrm{Sens}_d \le \|\widetilde{B}\|_2 \Gamma_{\mathrm{fb}}. \tag{21}$$

For a scalar decision, write $J = \chi$ and assume $B \neq 0$. Then

$$\mathrm{Sens}_d = \frac{\|BW^{-1/2}\|_2}{|1 - \chi|}, \qquad \|D_s q^\star - A_s\|_{W \to W} = \frac{\|W^{1/2} A_d\|_2 \|BW^{-1/2}\|_2}{|1 - \chi|}. \tag{22}$$

Equation (20) remains exact without the full-row-rank condition; only the lower bound in (21) can fail because the most amplified decision direction may be unreachable through the direct score response. This distinction prevents Γ_{fb} from being interpreted as an observable score-to-decision gain in directions that cannot be excited by score perturbations. Theorem 4 applies only on a regular stratum; active-set switches are handled by Theorem 5.

The derivative of repeated repair is J; $\rho(J)$ controls local fixed-point dynamics, whereas Γ_{fb} controls equilibrium amplification. In the positive scalar regime they reduce to the same critical parameter χ, but they can separate under nonnormal vector feedback; Appendix C.15 gives an exact realizable construction rather than claiming real-data validation of the vector case.

6.2 Smooth capacity decision

For (7), implicit differentiation gives

$$D\Theta_\tau(q)[h] = \frac{\sum_i w_i \sigma_i(1-\sigma_i)h_i}{\sum_i w_i \sigma_i(1-\sigma_i)}, \qquad \sigma_i = \sigma\left(\frac{q_i - \Theta_\tau(q)}{\tau}\right). \tag{23}$$

Thus H is a normalized nonnegative weighting concentrated near the decision boundary. In the scalar experiments, $\chi = HA_d$ is evaluated analytically or checked by finite differences.

7 Critical Feedback beyond a Fixed Active Set

Regular-stratum derivatives do not cover active-set switches. Fix (s, d), let $q = P(s, d)$, and let $I(q, d)$ contain all active rows. Let $\mathcal{D}_{\text{feas}} := \{d' : \mathcal{K}(d') \neq \varnothing\}$. Hadamard limits below are taken relative to the feasible parameter domain. We call (h, η) *admissible* if there exist $t_n \downarrow 0$ and $\eta_n \to \eta$ with $d + t_n\eta_n \in \mathcal{D}_{\text{feas}}$ and if the first-order feasible set

$$\begin{aligned} \mathcal{T}(\eta) &= \{v : C_I v \leq S_I \eta\}, \qquad \beta(\eta) = \min_{v \in \mathcal{T}(\eta)} \langle q - s, v\rangle_W, \\ \mathcal{C}(\eta) &= \{v \in \mathcal{T}(\eta) : \langle q - s, v\rangle_W = \beta(\eta)\} \end{aligned} \tag{24}$$

is nonempty. For such a direction, $\beta(\eta)$ is finite and attained: if $\lambda_I \geq 0$ is an active KKT multiplier, then $\langle q - s, v\rangle_W = -\lambda_I^\top C_I v \geq -\lambda_I^\top S_I \eta$ on $\mathcal{T}(\eta)$. Hence $\mathcal{C}(\eta) \neq \varnothing$. For $\eta = 0$, $\mathcal{C}(0)$ is the usual critical cone.

Theorem 5 (Directional best response and equilibrium at regime changes)**.** *For the strongly convex polyhedral projection (1), $(s, d) \mapsto P(s, d)$ is locally Lipschitz and piecewise affine on its feasible parameter domain. For every admissible direction (h, η), its Hadamard directional derivative is the unique solution*

$$P'((s, d); (h, \eta)) = \arg\min_{v \in \mathcal{C}(\eta)} \frac{1}{2}\|v - h\|_W^2. \tag{25}$$

No strict complementarity or fixed active set is required. For the equilibrium branch, suppose in addition that Θ is differentiable at an SCWR equilibrium and that there exist neighborhoods $\mathcal{S}$ of s and $\mathcal{U}$ of $d^\star$ such that, for every $\tilde{s} \in \mathcal{S}$, the map $T_{\tilde{s}}(d) := \Theta(P(\tilde{s}, d))$ maps $\mathcal{U}$ into itself and is γ-contractive there for one common $\gamma < 1$. Then the equilibrium branch is Hadamard directionally differentiable, with

$$v_h := q^{\star\prime}(s; h) = P'((s, d^\star); (h, \delta_h)), \qquad \delta_h := d^{\star\prime}(s; h) = D\Theta(q^\star)[v_h]. \tag{26}$$

The uniform local contraction is an additional sufficient condition, not an automatic consequence of polyhedrality. A uniform version of the global $\kappa < 1$ bound in Theorem 3 over $\tilde{s} \in \mathcal{S}$ is a sufficient special case. Appendix A.9 gives a two-stage liminf/recovery proof of (25), making explicit why the first-order critical set is selected before the strongly convex second-order projection.

8 Algorithm and Computational Structure

For fixed d, (1) is projection onto groupwise monotone cones, boxes, and a small number of halfspaces. For a current halfspace $H(d) = \{q : a^\top q \leq b(d)\}$, a denotes its normal and $b(d)$ its (possibly decision-dependent) right-hand side. Weighted Dykstra iterations alternate weighted PAVA, clipping, and the weighted halfspace projection

$$q \leftarrow q - \frac{[a^\top q - b(d)]_+}{a^\top W^{-1} a} W^{-1} a. \tag{27}$$

For scalar d, we solve $\Psi_s(d) := \Theta(P(s, d)) = d$ by fixed-point iteration in weak feedback or a safeguarded scalar root method near criticality. Sorting is done once; weighted PAVA and each halfspace pass are $O(N)$. Appendix B gives pseudocode and stopping criteria.

9 Experiments

The experiments test four questions: whether re-decision creates the residual predicted by Theorem 1; whether closing the loop restores smooth-decision feasibility; whether the scalar resolvent predicts perturbation amplification; and where the batch smooth-equilibrium claims stop.

Table 2: Real-data deployment results under calibration-selected feedback-active policies. SI is the fraction of one-shot runs that become infeasible after re-decision. SCWR cost is the probability-normalized groupwise squared transport cost in (5); shown SCWR rows converge to solver tolerance.

Dataset	Seeds	SI (%)	One-shot viol.	SCWR viol.	SCWR cost / ΔAUC
Adult	20	100	9.97×10^{-4}	3.63×10^{-10}	$0.0112 / -9.15\times10^{-4}$
Bank	20	100	3.32×10^{-3}	6.09×10^{-10}	$0.0086 / -1.78\times10^{-4}$
German Credit	50	94	8.02×10^{-3}	4.79×10^{-10}	$0.0259 / -7.11\times10^{-4}$
CDC Diabetes	10	100	2.11×10^{-3}	5.64×10^{-10}	$0.0050 / +1.45\times10^{-5}$
ACS-CA	10	100	3.57×10^{-3}	5.22×10^{-10}	$0.0093 / -9.90\times10^{-5}$
CA→NY	10	100	1.07×10^{-2}	7.04×10^{-10}	$0.0238 / -2.62\times10^{-5}$
CA→TX	10	100	9.60×10^{-3}	7.04×10^{-10}	$0.0167 / -2.57\times10^{-4}$

9.1 Protocol

We use Adult (Becker and Kohavi, 1996), Bank Marketing (Moro et al., 2014), South German Credit (Grömping, 2019), Heart Disease (Janosi et al., 1989), CDC Diabetes (CDC, 2023), and ACSIncome (Ding et al., 2021). The repair study starts from fixed held-out score vectors; SCWR does not retrain the predictor, and all repair variants operate on the same scores within a run. The score-generation/preprocessing configurations are frozen in the supplementary experiment records. Policies are selected only on calibration data from a prespecified feedback-active grid, frozen before test evaluation, and verified by a 140/140 hash audit (Appendix C). The self-invalidation frequencies below are therefore conditional on this feedback-active policy family; they are not population prevalence estimates for arbitrary deployed repairs. ACS transfer freezes the California policy for New York or Texas. Feasibility uses the smooth map (7); all failed outer runs are retained.

Implementation. All experiments were run in CPU mode on Windows 11 (10.0.22631, 64-bit) with Python 3.9.13, NumPy 1.21.5, SciPy 1.9.1, Pandas 1.4.4, and Matplotlib 3.5.2. Solver comparisons use the same software stack and score representation within each audit, so the reported timing differences do not rely on GPU acceleration or a change of computational backend.

9.2 One-shot failure and deployment-consistent repair

One-shot repair becomes infeasible after re-decision in 94–100% of the runs in Table 2. On converged runs, SCWR drives the smooth deployed residual to numerical tolerance while keeping $|\Delta\mathrm{AUC}| < 10^{-3}$. More importantly than the failure frequency, the counterfactual residual from Theorem 1 matches the stored deployment residual to floating-point precision. Under CA→NY and CA→TX transfer, mean one-shot violations increase to 1.07×10^{-2} and 9.60×10^{-3}, while re-solving SCWR on the deployment batch remains feasible under the smooth rule.

The stricter E0 audit has only 25/50 converged Heart runs. Throughout the paper, this means that the prespecified numerical protocol did not meet its stopping criterion within the recorded budget; it is not interpreted as evidence that an equilibrium does not exist. The existence theorem applies under Assumption 1, whereas E0 does not certify those conditions run-by-run. We therefore use Heart primarily as a stress test for near-critical feedback rather than as a counterexample to existence.

Why not directly solve the joint self-consistent program? A 20-bin joint nonlinear baseline optimizes repaired quantiles and the decision simultaneously. Its best-found feasible points have 4.8%, 25.7%, and 74.9% lower transport cost on Adult, German Credit, and paired Heart runs, but require roughly $625\times$, $1273\times$, and $4.5\times$ the runtime. Both methods in this comparison are evaluated on the same 20-bin representation, so these costs are not numerically comparable to the full-resolution cost column in Table 2. This does not make SCWR an approximation to the same optimizer: the joint program is a normative redesign of repair and decision as one nonconvex problem, whereas SCWR is the stationary state implied by a modular best-response repair followed by re-decision. Appendix C reports the paired comparison.

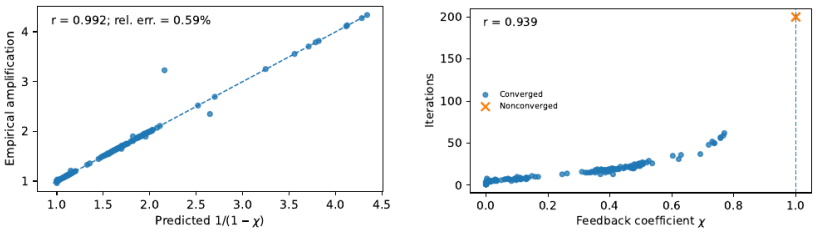

Figure 2: Scalar feedback validation. Left: predicted versus empirical amplification over 169 stable trials. Right: feedback coefficient versus fixed-point iterations; crosses mark the two nonconvergent critical Heart splits.

9.3 Feedback amplification and computational criticality

For a perturbed score vector $\tilde{s}$, define

$$A_{\mathrm{emp}} := \frac{|d^\star(\tilde{s}) - d^\star(s)|}{|T_{\tilde{s}}(d^\star(s)) - d^\star(s)|}, \qquad A_{\mathrm{th}} := \frac{1}{|1-\chi|}, \tag{28}$$

when the denominator is nonzero. This normalization removes the direct score-to-decision factor and isolates feedback amplification. Across 169 stable trials, the mean relative error between A_{th} and A_{emp} is 0.59%, with pooled Pearson correlation $r = 0.992$. We use the pooled correlation only as a descriptive mechanism diagnostic, not as an independence-based significance test. The feedback coefficient also tracks fixed-point iteration count ($r = 0.939$); two diagnostic Heart splits with $\chi \approx 1$ hit the 200-iteration cap.

9.4 Boundary and robustness audits

The stored audits make the scope of the claims explicit. Table 3 reports the two deployment boundaries that matter most. A calibration-frozen SCWR map improves held-out violation on Bank and ACS-CA/NY, but does not reproduce the near-zero residual obtained by re-solving the batch equilibrium; Adult and ACS-TX paired intervals include zero. Low-temperature smooth and weighted hard selected sets overlap strongly, but hard-decision residuals remain nonzero. Thus neither inductive freezing nor hard thresholding is silently treated as covered by the smooth batch theorem.

Table 3: Deployment-boundary audits. Panel A applies calibration-frozen maps to unseen scores; CI is SCWR-map minus one-shot-map violation. Panel B compares the smooth selected set with weighted hard top-k at $\tau = 0.02$.

A. Frozen-map inductive deployment				
Target	One-shot	SCWR map	Test-side SCWR	95% CI
Adult	.00901	.00876	4.0×10^{-10}	$[-5.17, .26] \times 10^{-4}$
Bank	.02666	.02470	6.8×10^{-10}	$[-2.38, -1.42] \times 10^{-3}$
ACS-CA	.00787	.00608	6.5×10^{-10}	$[-2.30, -.87] \times 10^{-3}$
ACS-NY	.01792	.01591	7.2×10^{-10}	$[-2.12, -1.87] \times 10^{-3}$
ACS-TX	.00221	.00191	4.6×10^{-10}	$[-1.39, .86] \times 10^{-3}$

B. Smooth versus hard capacity decision			
Dataset	Jaccard	Smooth viol.	Hard viol.
Adult	.9974	3.63×10^{-10}	1.34×10^{-3}
German Credit	.9838	4.10×10^{-10}	2.82×10^{-3}
ACS-CA	.9952	6.55×10^{-10}	3.45×10^{-4}

9.5 Derivative and solver audits

Table 4 directly checks the regular-stratum derivative in Theorem 4 against a two-sided full SCWR re-solve at $\epsilon = 10^{-4}$. All switching and nonconvergent records remain in the archive; the pass rate is evaluated on the regular-stratum stable subset.

Table 4: Regular-stratum implicit-gradient audit. "Stable" denotes records for which the ordinary Jacobian is applicable; Pass counts relative error below 1%.

Dataset	Stable / records	Pass $< 1\%$	Median rel. error
Adult	20/20	20/20	1.12×10^{-4}
Bank	20/20	19/20	1.99×10^{-7}
Heart	25/50	25/25	9.91×10^{-9}
German Credit	50/50	50/50	9.96×10^{-9}

The single Bank miss at 10^{-4} falls below 1% at a coarser step in the prespecified sweep, while random-sign stress directions frequently switch active sets, consistent with the need for Theorem 5.

Solver and stress-test audit. Table 5 collects the stored CPU checks that complement the derivative audit. The three outer solvers reach the same recorded CDC-50k transport cost and equilibrium tolerance while differing markedly in runtime. The remaining rows stress sample size, group imbalance, and the number of moving constraints without changing the main protocol. These checks target numerical stability and scaling rather than introduce a second performance objective.

Table 5: CPU solver and robustness audits from stored CDC runs. For grouped settings, every listed configuration attains the reported convergence count. "Inner time" refers to the mean inner-projection time.

Audit	Settings	Recorded outcome	Timing
Outer solver	Brent	cost 0.450082; tol. met	22.89 s
	FP, damping 0.5	cost 0.450082; tol. met	0.90 s
	FP, damping 1.0	cost 0.450082; tol. met	0.38 s
Sample size	5k / 25k / 50k	5/5 each; viol. $[5.5, 6.1] \times 10^{-10}$	3.92–8.85 s
Group fraction	.50 / .70 / .90 / .95	5/5 each; viol. $< 7 \times 10^{-10}$	8.31–8.58 s
Moving constr.	1 / 2 / 3	5/5 each	inner time 0.11 / 0.22 / 0.41 s

Taken together, Tables 2–5 separate the empirical roles of the proposed checks. Table 2 establishes the deployment-consistency gap and its correction; Table 3 tests nearby deployment variants; Table 4 verifies the regular-stratum derivative; and Table 5 shows solver agreement and stable convergence/residual behavior across controlled CDC stress settings. This organization mirrors the theoretical decomposition into feasibility, sensitivity, and numerical solution rather than treating all experiments as interchangeable accuracy benchmarks. Appendix C reports the full audits, uncertainty summaries, and retained failures.

10 Conclusion

A frozen score repair can be exact for the decision used to compute it yet infeasible after the repaired scores trigger a new decision. The resulting difficulty is endogenous feasibility, not inaccurate solution of the frozen projection. SCWR closes this loop by requiring the minimum-distortion repair at the equilibrium decision to reproduce that decision. The theory separates existence, global contraction, local dynamics, and intrinsic equilibrium sensitivity; in the scalar experiments, the $1/(1-\chi)$ resolvent law predicts perturbation amplification and the slowdown near critical feedback. The critical-cone analysis extends the first-order description across active-set changes. Direct-joint and boundary audits clarify the solution concept: joint redesign can reduce transport cost, while frozen-map transfer and hard selection approximate the batch smooth equilibrium. Together with the gradient and solver audits, they separate feedback-induced feasibility from numerical solution effects and support broader deployment-consistent repair.

LIMITATIONS

Our analysis is developed for the batch decision setting used in the main experiments, with grouped one-dimensional scores, polyhedral repair constraints, and the smooth capacity map. The appendix also studies calibration-frozen maps, low-temperature hard selection, and a realizable vector-feedback construction, providing direct checks of nearby variants. Extending the exact guarantees to broader inductive and multidimensional decision rules is a natural next step.

AI USE STATEMENT

Generative AI tools were used only for language polishing and limited auxiliary support in experimental design and result analysis. They did not generate or infer any experimental result. The authors independently checked the conceptual and theoretical contributions, theorem statements, proofs, code, experiments, figures, and cited sources, and take full responsibility for the final manuscript.

ETHICS STATEMENT

The experiments use public benchmark datasets to study constrained score post-processing. Results involving demographic groups should not be interpreted as deployment recommendations for eligibility, medical treatment, or credit decisions. The Heart Disease and CDC experiments are methodological stress tests, not clinical claims. Any real deployment would require domain-specific policy design, causal and fairness analysis, and separate validation of the chosen constraints.

REPRODUCIBILITY STATEMENT

Appendix B gives solver pseudocode and tolerances, Appendix A contains complete proofs, and Appendix C records datasets, splits, calibration, hash audits, direct-joint and inductive baselines, smooth/hard boundary tests, gradient audits, robustness sweeps, uncertainty summaries, timing, and failed runs. The supplementary package contains the CPU implementation, experiment configurations, seed-level outputs, and aggregation scripts.

REFERENCES

Brandon Amos and J. Zico Kolter. OptNet: Differentiable optimization as a layer in neural networks. In *ICML*, 2017.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *NeurIPS*, 2019.

Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2nd edition, 2017.

Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-Lopez, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. In *NeurIPS*, 2022.

James P. Boyle and Richard L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in Order Restricted Statistical Inference*, 1986.

Centers for Disease Control and Prevention. CDC Diabetes Health Indicators. UCI Machine Learning Repository, 2023.

Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring Adult: New datasets for fair machine learning. In *NeurIPS*, 2021.

Ulrike Grömping. South German Credit. UCI Machine Learning Repository, 2019.

Florian Gunsilius, Meng Hsuan Hsieh, and Myung Jin Lee. Tangential Wasserstein projections. *JMLR*, 25(69):1–41, 2024.

Marouane Il Idrissi, Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, and Jean-Michel Loubes. Quantile-constrained Wasserstein projections for robust interpretability of numerical and machine learning models. *Electronic Journal of Statistics*, 18(2):2721–2770, 2024.

Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano. Heart Disease. UCI Machine Learning Repository, 1989.

Jakwang Kim, Young-Heon Kim, and Chan Park. Extension of coupling via the projection of optimal transport. arXiv:2603.27535, 2026.

Takeru Matsuda and Ting-Kam Leonard Wong. Shape-constrained density estimation with Wasserstein projection. arXiv:2603.08939, 2026.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. In *NeurIPS*, 2020.

Sergio Moro, Paulo Cortez, and Paulo Rita. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62:22–31, 2014.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *ICML*, 2020.

Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport*. Now Publishers, 2019.

Tim Robertson, F. T. Wright, and Richard L. Dykstra. *Order Restricted Statistical Inference*. Wiley, 1988.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer, 1998.

Alexander Shapiro. Differentiability properties of metric projections onto convex sets. *Journal of Optimization Theory and Applications*, 169(3):953–964, 2016.

Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.

Zifan Wang, Changxin Liu, Thomas Parisini, Michael M. Zavlanos, and Karl H. Johansson. Constrained optimization with decision-dependent distributions. *IEEE Transactions on Automatic Control*, 70(8):5176–5189, 2025.

Siyi Wang, Zifan Wang, and Karl H. Johansson. Wasserstein distributionally robust performative prediction. arXiv:2602.06730, 2026.

Guangzheng Zhong, Yang Liu, and Jiming Liu. Towards performatively stable equilibria in decision-dependent games for arbitrary data distribution maps. *Machine Learning*, 115:197, 2026.

A Complete Proofs

A.1 Proof of Proposition 1

For group g, μ_g and $\nu_g(q_g)$ place the same masses $\omega_{g1}, \dots, \omega_{gn_g}$ on two nondecreasing support sequences. Let $a_{g0} = 0$ and $a_{gi} = \sum_{k \le i} \omega_{gk}$. On $[a_{g,i-1}, a_{gi})$, the quantile functions equal s_{gi} and q_{gi}, respectively. Hence

$$W_2^2(\mu_g, \nu_g) = \int_0^1 |Q_{\mu_g}(u) - Q_{\nu_g}(u)|^2 \, du = \sum_i \omega_{gi}(q_{gi} - s_{gi})^2.$$

Multiplying by π_g, summing over g, and using $w_{gi} = \pi_g \omega_{gi}$ gives (5). The same quantile representation proves optimality of the monotone rank-preserving coupling. □

A.2 Proof of Theorem 1

By definition,

$$c_j(q_1, d_1) - c_j(q_1, d_0) = -S_j d_1 + S_j d_0 = S_j(d_0 - d_1),$$

which proves (9). If the row is active at (q_1, d_0), then $c_j(q_1, d_0) = 0$; taking the positive part gives (10). No approximation or regularity assumption is used. □

A.3 Proof of Theorem 2

Fix s. For each $d \in \mathcal{D}$, the objective $q \mapsto \frac{1}{2}\|q - s\|_W^2$ is continuous, coercive, and strictly convex because $W \succ 0$. Minimization over the nonempty closed convex set $\mathcal{K}(d)$ therefore has a unique solution $P(s, d)$.

We next prove continuity of the best response. Let $d_n \to d$ and $q_n = P(s, d_n)$. Assumption 1 makes $\{q_n\}$ bounded, so every subsequence has a convergent further subsequence $q_n \to \bar{q}$. Outer semicontinuity gives $\bar{q} \in \mathcal{K}(d)$. For any $z \in \mathcal{K}(d)$, inner semicontinuity provides $z_n \in \mathcal{K}(d_n)$ with $z_n \to z$. Optimality yields $\|q_n - s\|_W^2 \le \|z_n - s\|_W^2$. Passing to the limit shows that $\bar{q}$ minimizes over $\mathcal{K}(d)$; uniqueness gives $\bar{q} = P(s, d)$. Thus $P(s, \cdot)$ is continuous.

The map $T_s(d) = \Theta(P(s, d))$ is continuous and maps the nonempty compact convex set $\mathcal{D}$ into itself. Brouwer's theorem gives $d^\star = T_s(d^\star)$; setting $q^\star = P(s, d^\star)$ yields (2). Finally, metric projection onto a closed convex set satisfies

$$q = P(s, d) \iff q \in \mathcal{K}(d), \quad \langle q - s, z - q\rangle_W \ge 0 \; \forall z \in \mathcal{K}(d),$$

and substituting $d = \Theta(q)$ gives (8) and its converse. □

A.4 Proof of Proposition 2

The KKT system for (1) is

$$W(q - s) + C^\top \lambda = 0, \qquad Cq - r - Sd \le 0, \quad \lambda \ge 0, \quad \lambda_i(C_i q - r_i - S_i d) = 0.$$

Because $W \succ 0$, the primal solution is unique. Fix a linearly independent active basis I. On the parameter region where this basis supports the solution, active equalities and stationarity give

$$q = s - W^{-1}C_I^\top (C_I W^{-1} C_I^\top)^{-1}(C_I s - r_I - S_I d),$$

which is affine in d. The associated primal-feasibility, dual-nonnegativity, and inactivity conditions are linear inequalities in d, hence the cell is polyhedral. There are finitely many bases, so the unique solution map is continuous piecewise affine. Along any segment between d and d', the path crosses finitely many cells and the derivative on a cell is

$$P_{d,I} = W^{-1}C_I^\top (C_I W^{-1} C_I^\top)^{-1} S_I.$$

In W-norm its Euclidean operator norm is the term in (12). Integrating the piecewise-constant derivative along the segment gives the claimed global bound. Degenerate boundaries are covered by taking the maximum over all realized local affine selections. □

A.5 Proof of Theorem 3

The two Lipschitz inequalities in (11) imply

$$\|T_s(d) - T_s(d')\|_2 \le L_\theta \|P(s, d) - P(s, d')\|_W \le \kappa \|d - d'\|_2.$$

Thus T_s is a contraction when $\kappa < 1$, and Banach's theorem gives uniqueness and (13).

For perturbation stability, let $q^\star = P(s, d^\star)$ and $\tilde{q}^\star = P(\tilde{s}, \tilde{d}^\star)$. Because W, C, r, S are fixed, projection onto the fixed set $\mathcal{K}(\tilde{d}^\star)$ is nonexpansive in its center under the same W-metric:

$$\|\tilde{q}^\star - q^\star\|_W \le \|P(\tilde{s}, \tilde{d}^\star) - P(s, \tilde{d}^\star)\|_W + \|P(s, \tilde{d}^\star) - P(s, d^\star)\|_W \le \|\tilde{s} - s\|_W + L_P \|\tilde{d}^\star - d^\star\|_2.$$

Since $\tilde{d}^\star = \Theta(\tilde{q}^\star)$ and $d^\star = \Theta(q^\star)$,

$$\|\tilde{d}^\star - d^\star\|_2 \le L_\theta \|\tilde{q}^\star - q^\star\|_W.$$

Combining yields (14). For (15), write $d_1 = T_s(d_0)$ and $d^\star = T_s(d^\star)$:

$$\|d^\star - d_0\|_2 \le \|d^\star - d_1\|_2 + \|d_1 - d_0\|_2 \le \kappa \|d^\star - d_0\|_2 + \|d_1 - d_0\|_2.$$

The score bound then follows from $\|P(s, d^\star) - P(s, d_0)\|_W \le L_P \|d^\star - d_0\|_2$. □

A.6 Proof of Proposition 3

On a regular stratum, active inequalities remain equalities and inactive ones stay strictly slack. The local problem is equality-constrained:

$$\min_q \ \tfrac{1}{2}(q-s)^\top W(q-s) \quad \text{s.t.} \quad C_I q = r_I + S_I d.$$

Stationarity gives $W(q-s) + C_I^\top \lambda = 0$, hence $q = s - W^{-1}C_I^\top \lambda$. Substitution yields $\lambda = M_I^{-1}(C_I s - r_I - S_I d)$ because $M_I \succ 0$. Substituting back proves (17); differentiating gives (18). □

A.7 Proof of Theorem 4

Define $F(s,d) = d - \Theta(P(s,d))$. At $(s,d^\star)$, $F(s,d^\star) = 0$. Proposition 3 gives

$$D_d F = I - HA_d = I - J, \qquad D_s F = -HA_s = -B.$$

If $I - J$ is nonsingular, the implicit-function theorem gives a differentiable local branch and

$$D_s d^\star = -(D_d F)^{-1} D_s F = (I-J)^{-1}B.$$

Differentiating $q^\star(s) = P(s, d^\star(s))$ gives the second identity in (19). For $z = W^{1/2}h$, $\|h\|_W = \|z\|_2$ and

$$D_s d^\star[h] = (I-J)^{-1}BW^{-1/2}z,$$

which proves (20). The upper bound in (21) is submultiplicativity. For the lower bound, let $R = (I-J)^{-1}$ and choose unit y with $\|Ry\|_2 = \|R\|_2$. Full row rank of $\widetilde{B}$ makes it surjective, so $x = \widetilde{B}^\dagger y$ satisfies $\widetilde{B}x = y$ and $\|x\|_2 \le 1/\sigma_{\min}(\widetilde{B})$. Normalizing x gives

$$\|R\widetilde{B}\|_2 \ge \sigma_{\min}(\widetilde{B})\|R\|_2.$$

For $p = 1$, $R = (1-\chi)^{-1}$, proving the first formula in (22). The second follows from $D_s q^\star - A_s = A_d R B$ and the spectral norm of a rank-one matrix. □

Global-to-local relation. In the compatible norms, $\|A_d\|_{2\to W} \le L_P$ and $\|H\|_{W\to 2} \le L_\theta$, so $\|J\|_2 \le \kappa$ and $\rho(J) \le \kappa$. If $\kappa < 1$, the Neumann series gives $\Gamma_{\mathrm{fb}} \le (1-\kappa)^{-1}$.

A.8 Smooth capacity derivative

Let $G(q,d) = \sum_i w_i \sigma((q_i - d)/\tau) - \alpha$. Then $\partial_d G = -\tau^{-1}\sum_i w_i \sigma_i(1-\sigma_i) < 0$, so the root is unique. Its limits at $d \to \pm\infty$ have opposite signs. At the root $d = \Theta_\tau(q)$, differentiating $G(q, \Theta_\tau(q)) = 0$ in direction h gives (23).

A.9 Proof of Theorem 5

The strongly convex quadratic program (1) has a fixed positive-definite Hessian and polyhedral constraints whose right-hand side is affine in d. Its unique solution map is therefore continuous piecewise affine on the feasible parameter domain; in particular it is locally Lipschitz and Hadamard directionally differentiable (Rockafellar and Wets, 1998; Shapiro, 2016). We identify the derivative explicitly.

Fix an admissible direction (h, η) and arbitrary feasible-domain sequences $t_n \downarrow 0$, $h_n \to h$, $\eta_n \to \eta$ with $d_n := d + t_n\eta_n \in \mathcal{D}_{\mathrm{feas}}$. Let

$$s_n = s + t_n h_n, \qquad q_n = P(s_n, d_n), \qquad v_n = (q_n - q)/t_n.$$

Local Lipschitz continuity makes $\{v_n\}$ bounded. Consider a convergent subsequence $v_n \to v$. For every active row, feasibility gives $C_i q_n \le r_i + S_i d_n$ and $C_i q = r_i + S_i d$, hence $C_i v \le S_i \eta$ in the limit. Inactive rows remain inactive for small t_n. Thus $v \in \mathcal{T}(\eta)$. For the full active-row matrix C_I, the KKT conditions provide a multiplier $\lambda_I \ge 0$ satisfying $W(q-s) + C_I^\top \lambda_I = 0$. Therefore $\langle q - s, u\rangle_W = -\lambda_I^\top C_I u \ge -\lambda_I^\top S_I \eta$ for every $u \in \mathcal{T}(\eta)$. Since $\mathcal{T}(\eta)$ is a nonempty closed polyhedron, the linear minimum is finite and attained, so $\mathcal{C}(\eta)$ is nonempty.

Conversely, for any $u \in \mathcal{T}(\eta)$, define $\mathcal{T}_n = \{z : C_I z \le S_I \eta_n\}$. These sets are nonempty for large n because $v_n \in \mathcal{T}_n$. A standard Hoffman bound for the fixed normal matrix C_I and convergent right-hand sides then provides $u_n \in \mathcal{T}_n$ with $u_n \to u$. Because inactive rows have strictly positive slack at the base point, $q + t_n u_n \in \mathcal{K}(d_n)$ for all large n. Hence every $u \in \mathcal{T}(\eta)$ has a feasible recovery sequence.

Now compare objective values. Up to terms independent of the candidate direction,

$$\frac{1}{2}\|q + t_n v - (s + t_n h_n)\|_W^2 = t_n \langle q - s, v\rangle_W + \frac{t_n^2}{2}\|v - h_n\|_W^2 + \text{const.} \tag{29}$$

We make the two stages explicit. First, let $\hat{u} \in \mathcal{T}(\eta)$ satisfy $\langle q - s, \hat{u}\rangle_W < \langle q - s, v\rangle_W$. A recovery sequence $\hat{u}_n \to \hat{u}$ would improve the $O(t_n)$ term in (29) by a fixed multiple of t_n, while the quadratic difference is only $O(t_n^2)$, contradicting optimality of q_n for large n. Therefore every cluster point v minimizes the linear term over $\mathcal{T}(\eta)$, i.e. $v \in \mathcal{C}(\eta)$.

Second, restrict to $u \in \mathcal{C}(\eta)$. The $O(t_n)$ terms of v_n and any recovery sequence $u_n \to u$ agree asymptotically. Dividing the optimality inequality by t_n^2 and taking the liminf/limsup using the recovery sequence yields

$$\frac{1}{2}\|v - h\|_W^2 \le \frac{1}{2}\|u - h\|_W^2 \qquad \forall u \in \mathcal{C}(\eta).$$

Because the objective is strongly convex on the closed convex set $\mathcal{C}(\eta)$, this minimizer is unique. Hence every cluster point of v_n is the same vector, proving the Hadamard derivative (25).

For the equilibrium branch, let $s_n = s + t_n h_n$ and let $d_n^\star$ be the unique fixed point in the common neighborhood. Uniform contraction gives

$$(1 - \gamma)\|d_n^\star - d^\star\|_2 \le \|T_{s_n}(d^\star) - T_s(d^\star)\|_2 = O(t_n),$$

so $\delta_n = (d_n^\star - d^\star)/t_n$ is bounded. Along a convergent subsequence $\delta_n \to \delta$, Hadamard differentiability of P and differentiability of Θ give

$$v = P'((s, d^\star); (h, \delta)), \qquad \delta = D\Theta(q^\star)[v].$$

For fixed h, define $R_h(\delta) = D\Theta(q^\star)[P'((s, d^\star); (h, \delta))]$. Passing the finite-$t$ contraction inequality to the directional limit gives $\|R_h(\delta_1) - R_h(\delta_2)\|_2 \le \gamma\|\delta_1 - \delta_2\|_2$. Thus R_h has a unique fixed point δ_h; all cluster points coincide, proving (26). □

B Algorithmic Details

For fixed d, Dykstra's algorithm alternates weighted projections onto (i) groupwise monotone cones by weighted PAVA, (ii) the box $[0, 1]^N$ for probabilistic scores, and (iii) each moving or fixed halfspace using (27). The Dykstra correction vectors are retained; cyclic projections without them need not return the metric projection onto the intersection.

Outer equilibrium. For scalar decisions, the implementation provides raw fixed-point updates $d_{t+1} = \Psi_s(d_t)$ and a safeguarded root mode for $\Psi_s(d) - d = 0$. Each run records the decision residual, maximum deployed constraint residual, outer iterations, and an analytic or finite-difference estimate of χ.

Pseudocode. (1) Sort test scores groupwise and construct normalized probability weights. (2) Given d, compute $q = P(s, d)$ with weighted Dykstra/PAVA until both projection change and maximum constraint residual are below tolerance. (3) Compute $\Psi_s(d) = \Theta(q)$. (4) Update d by fixed-point iteration, damping, or the safeguarded root method. (5) Stop when both $|\Psi_s(d) - d|$ and the deployed constraint residual meet tolerance, then map repaired ordered scores back to original sample indices.

C Additional Experimental Details and Results

C.1 Data, splits, and policy selection

Adult, Bank Marketing, South German Credit, Heart Disease, and CDC Diabetes are taken from the UCI repository; ACSIncome is generated with Folktables. For confirmatory experiments, policy

parameters are selected only on calibration data from a fixed candidate grid and then frozen. The final test split is not used to choose margins or feedback coupling. Evaluation records raw utility, frozen-decision feasibility, the recomputed decision, deployed one-shot violation, the SCWR equilibrium, normalized groupwise Wasserstein distortion, outer iterations, and convergence status. All seed-level outputs are retained.

C.2 E0: INDEPENDENT NO-TEST-TUNING AUDIT

The stricter train–calibration–test protocol serializes and SHA-256 hashes every selected policy before test evaluation. All 140 policy audits pass. Audit success means the pre-test policy artifact is unchanged; it does not imply outer-solver convergence. The Heart records are retained as a critical

Table 6: E0 calibration-frozen audit.

Dataset	Audit pass	Outer convergence
Adult	20/20	20/20
Bank	20/20	20/20
German Credit	50/50	50/50
Heart	50/50	25/50

stress test. The 25 nonconvergent E0 runs are numerical protocol failures under the recorded stopping rule, not demonstrations of equilibrium nonexistence.

C.3 E1: DIRECT JOINT SELF-CONSISTENT OPTIMIZATION

E1 solves a nonlinear program over repaired quantiles and the decision on a 20-bin discretization using multiple starts with SLSQP and trust-constr. We report the best-found feasible point and do not treat it as a certified global optimum. The direct program reaches negligible feasibility residual

Table 7: E1 best-found joint self-consistent baseline on the common 20-bin representation. Gap is $(\mathrm{cost}_{\mathrm{SCWR}}/\mathrm{cost}_{\mathrm{joint}} - 1) \times 100\%$ on paired runs; these discretized costs are not numerically comparable to Table 2.

Dataset	Paired n	Joint cost	SCWR cost	Gap / runtime
Adult	20	0.0258	0.0271	$+4.8\%$ / $625\times$
German Credit	50	0.0271	0.0341	$+25.7\%$ / $1273\times$
Heart	49	0.0494	0.0863	$+74.9\%$ / $4.5\times$

and lower distortion. The comparison therefore clarifies the solution concept rather than establishing dominance: SCWR is a structured feedback equilibrium, not a certified global optimizer.

C.4 E2: FROZEN-MAP INDUCTIVE DEPLOYMENT

A calibration-set repair can be frozen as a monotone groupwise map and applied to unseen scores, but this is an inductive extension rather than the batch equilibrium analyzed in the theory. Its violations are therefore not numerically comparable to the batch one-shot values in Table 2 or the bootstrap summaries below. Frozen SCWR maps reduce violation on Bank and ACS-CA/NY; Adult and ACS-TX intervals include zero. Re-solving the batch equilibrium recovers numerical feasibility in every reported case. E2's historical stored W_2 column is omitted because those files used unnormalized survey weights.

C.5 E3: SMOOTH VERSUS HARD CAPACITY DECISIONS

At the prespecified $\tau = 0.02$, the smooth selected set strongly overlaps a weighted hard top-k prefix, but exact smooth feasibility does not imply hard feasibility. For ACS-CA, Jaccard increases from 0.9214 at $\tau = 0.10$ to 0.9986 at $\tau = 0.005$. We therefore use hard selection only as a boundary approximation.

Table 8: E2 held-out deployment. CI is the paired 95% bootstrap interval for SCWR-map minus one-shot-map violation.

Target	One-shot map	SCWR map	Test-side SCWR	95% CI
Adult	0.00901	0.00876	4.0×10^{-10}	$[-5.17, 0.26] \times 10^{-4}$
Bank	0.02666	0.02470	6.8×10^{-10}	$[-2.38, -1.42] \times 10^{-3}$
ACS-CA	0.00787	0.00608	6.5×10^{-10}	$[-2.30, -0.87] \times 10^{-3}$
ACS-NY	0.01792	0.01591	7.2×10^{-10}	$[-2.12, -1.87] \times 10^{-3}$
ACS-TX	0.00221	0.00191	4.6×10^{-10}	$[-1.39, 0.86] \times 10^{-3}$

Table 9: E3 soft/hard audit at $\tau = 0.02$.

Dataset	Jaccard	Soft violation	Hard violation
Adult	0.9974	3.63×10^{-10}	1.34×10^{-3}
German Credit	0.9838	4.10×10^{-10}	2.82×10^{-3}
ACS-CA	0.9952	6.55×10^{-10}	3.45×10^{-4}

C.6 E4: REGULAR-STRATUM IMPLICIT-GRADIENT AUDIT

E4 compares the regular-stratum implicit directional derivative of a weighted Brier loss with a full two-sided SCWR re-solve at $\epsilon = 10^{-4}$; no failed or switching record is deleted from the archive. The

Table 10: E4 regular-stratum gradient audit.

Dataset	Records	Stable	Pass $< 1\%$	Median rel. error
Adult	20	20	20	1.12×10^{-4}
Bank	20	20	19	1.99×10^{-7}
Heart	50	25	25	9.91×10^{-9}
German Credit	50	50	50	9.96×10^{-9}

single Bank miss at 10^{-4} falls below 1% at a coarser step in the prespecified sweep. Random-sign stress directions frequently switch active sets, motivating Theorem 5 rather than forcing an ordinary Jacobian through regime changes.

C.7 COUNTERFACTUAL IDENTITY CHECKS

For each diagnostic run, the archive stores d_0, q_1, d_1 and every moving-row residual. The prediction from (9) matches the observed residual to floating-point precision. Self-invalidation occurs in 20/20 Adult, 12/12 Bank diagnostics, 10/10 CDC, 10/10 ACS-CA, 9/10 ACS-NY, 8/10 ACS-TX, 10/10 CA→NY, 10/10 CA→TX, 47/50 German Credit, and 44/50 Heart runs. Where a run does not self-invalidate, the same identity matches the nonpositive deployed residual.

C.8 FEEDBACK SWEEPS

A perturbation trial is aggregated only when base and perturbed equilibrium runs converge and all recorded quantities are finite. The 169 stable trials comprise 20 Adult, 12 Bank, 50 German Credit, 10 CDC, 47 Heart, and 30 ACS state trials. The pooled correlation between $1/(1-\chi)$ and A_{emp} is 0.9924 and the mean relative error is 0.591%. We report this pooled correlation descriptively, not as an independence-based significance test. For fixed-point convergence, 170 converged base runs are available and the correlation between χ and outer iterations is 0.9388. Two excluded diagnostic Heart base runs have $\chi \approx 1$ and hit the 200-iteration cap.

C.9 HEART AS A FEEDBACK STRESS TEST

The controlled feedback sweep and E0 use different policy protocols. In the diagnostic sweep, 48/50 base runs converge; seeds 239 and 353 have feedback coefficient approximately one and hit the

200-iteration cap. Under the stricter calibration-frozen E0 family, 25/50 runs meet the recorded outer stopping rule. Both results are retained: the controlled sweep isolates the predicted critical mechanism, whereas E0 shows a substantially harder numerical regime under the confirmatory policy family.

C.10 Baselines and ablations

The supplementary result files contain raw scores, threshold-only adjustment, affine/group shifts, isotonic calibration, penalty-based post-processing, one-shot Wasserstein repair, and SCWR. Generic fixed-decision transforms are not designed to solve the same equilibrium problem, so we use them only to assess deployed feasibility after re-decision. Adult, Bank, CDC, German Credit, and ACS show the same qualitative pattern: one-shot Wasserstein repair is closer to deployed feasibility than generic fixed-decision transforms, while SCWR enforces the smooth equilibrium residual on converged runs. Penalty sweeps on Adult show that increasing a fixed-decision penalty does not remove the decision-feedback residual. We do not turn these heterogeneous methods into a leaderboard because their optimization targets differ.

C.11 Bootstrap uncertainty and efficiency

Paired bootstrap intervals summarize seed-to-seed uncertainty of one-shot violation. Adult has mean 9.97×10^{-4} with a 95% interval approximately $[5.84\times10^{-4}, 1.44\times10^{-3}]$; CA→NY has 1.07×10^{-2} with $[9.40\times10^{-3}, 1.20\times10^{-2}]$; CA→TX has 9.60×10^{-3} with $[7.26\times10^{-3}, 1.18\times10^{-2}]$. These are uncertainty summaries, not cross-domain effect-size estimates.

On CDC-50k, the conservative Brent configuration averages 22.89 seconds, fixed-point damping 0.5 averages 0.90 seconds, and unit-damping fixed-point updates average 0.38 seconds; all three reach the same recorded Wasserstein solution ($W_2 = 0.450082$) and satisfy the residual tolerance. Solver choice changes computational cost, not the local branch conditioning in Theorem 4.

C.12 Utility beyond ranking in the inductive audit

Table 11: Additional E2 utility diagnostics for the calibration-frozen SCWR map; changes are relative to raw test scores.

Target	ΔAUROC	ΔAUPRC	ΔBrier	ΔECE	Turnover
Adult	+0.000003	−0.000034	+0.000033	−0.000238	0.0031
Bank	−0.000124	−0.000738	−0.000088	−0.000993	0.0206
ACS-CA	+0.000089	+0.000254	−0.000084	−0.000346	0.0133
ACS-NY	+0.000133	+0.000362	−0.000062	+0.000589	0.0155
ACS-TX	+0.000004	+0.000057	+0.000182	+0.001446	0.0137

C.13 Sample-size, imbalance, and solver robustness

CDC pressure tests give 5/5 convergence at sample sizes 5k, 25k, and 50k, with mean smooth residuals between 5.5×10^{-10} and 6.1×10^{-10}. A single $n = 250$ smoke run does not meet the convergence flag and is retained rather than used as small-sample evidence. Under group fractions 0.50, 0.70, 0.90, and 0.95, all five seeds converge with mean residuals below 7×10^{-10}. Increasing moving constraints from one to three gives 5/5 convergence in each setting; mean inner-projection time rises from 0.11 to 0.22 to 0.41 seconds.

C.14 Controlled mechanism visualization

The synthetic mechanism check in the supplementary archive visualizes rank-preserving repair, the fixed-point map, one-shot self-invalidation, and the feedback coefficient as coupling changes. It is a mechanism visualization rather than primary empirical evidence.

C.15 A REALIZABLE NONNORMAL VECTOR-FEEDBACK EXAMPLE

The separation between $\rho(J)$ and Γ_{fb} occurs within the local SCWR model. Take $p = 2$, $N = 4$, $W = I_4$, $s = (1, 1, 0, 0)^\top$, and moving constraints $q_1 \le d_1$, $q_2 \le d_2$. At $d^\star = 0$, the active matrices are

$$C_I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \qquad S_I = I_2, \qquad r_I = 0,$$

so $q^\star = 0$; for $d_1, d_2 < 1$ this is a regular stratum. Equations (17)–(18) give

$$A_d = W^{-1} C_I^\top (C_I W^{-1} C_I^\top)^{-1} S_I = \begin{bmatrix} I_2 \\ 0 \end{bmatrix}, \qquad A_s = \mathrm{diag}(0, 0, 1, 1).$$

Choose the locally affine decision map $\Theta(q) = Hq$ with $H = [J_a \; I_2]$, where $J_a = \begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}$. Then $\Theta(q^\star) = d^\star$, $J = HA_d = J_a$, and $B = HA_s$ has full row rank through the last two score coordinates. We have $\rho(J) = 0$ for every a, but

$$(I - J)^{-1} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \qquad \Gamma_{\mathrm{fb}} = \sqrt{\frac{a^2 + 2 + |a|\sqrt{a^2 + 4}}{2}},$$

which grows as $|a|$.