

Online Resource 1: Mathematical supplement for “Auditing longitudinal data across institutional regimes: a graph-based compatibility protocol with evidence from Tunisia”

Anonymous Author^{1*}

^{1*} Anonymous for double-blind review.

Abstract

This supplement proves the formal results used in the article. It also gives positive and failure examples, the finite-sample level–shape calculation, a consistency extension for stationary ergodic edge defects, and the complete specification of the controlled simulation.

Keywords: graph consistency, convex projection, Farkas certificate, perturbation stability, ergodic consistency

1 Notation and standing conventions

Let $G = (V, E)$ be a finite connected graph with an arbitrary orientation. Write $n = |V|$ and $m = |E|$. Its incidence matrix $B \in \mathbb{R}^{m \times n}$ has row $e = (i, j)$ equal to $-e_i^\top + e_j^\top$. A positive diagonal matrix $W = \text{diag}(w_e : e \in E)$ defines the edge inner product

$$\langle u, v \rangle_W = \text{tr}(u^\top W v), \quad u, v \in \mathbb{R}^{m \times p}.$$

The constant vector in $\mathbb{R}^n$ is denoted by $\mathbf{1}$. Since G is connected, $\ker B = \text{span}\{\mathbf{1}\}$ and $B^\top W B$ is positive definite on $\mathbf{1}^\perp$.

For an edge window Q of cardinality N , let $H = (\mathbb{R}^p)^Q$ with

$$\langle r, s \rangle_H = \frac{1}{N} \sum_{q \in Q} r(q)^\top s(q), \quad \|r\|_H^2 = \frac{1}{N} \sum_{q \in Q} \|r(q)\|_2^2.$$

A set $C \subset \mathbb{R}^p$ is identified with the set of constant functions $q \mapsto c$, $c \in C$. Under this embedding, the article's edge residual is $D = p^{-1/2} \text{dist}_H(r, C)$.

2 Unique ridge fit

Proposition 2.1 (Unique regularised fit) *Let $X \in \mathbb{R}^{N \times (d+1)}$, $Y \in \mathbb{R}^{N \times p}$, $R = \text{diag}(0, 1, \dots, 1)$, and $\lambda > 0$. If the intercept column of X is non-zero, then*

$$\Phi(\Theta) = \|Y - X\Theta\|_F^2 + \lambda \|R\Theta\|_F^2$$

has the unique minimiser

$$\hat{\Theta} = (X^\top X + \lambda R^\top R)^{-1} X^\top Y.$$

Proof The Hessian of Φ is $2(X^\top X + \lambda R^\top R) \otimes I_p$. For $v = (v_0, \dots, v_d)^\top$,

$$v^\top (X^\top X + \lambda R^\top R) v = \|Xv\|_2^2 + \lambda \sum_{k=1}^d v_k^2.$$

If this quantity is zero, then $v_1 = \dots = v_d = 0$ and $Xv = v_0 \mathbf{1}_N = 0$. The non-zero intercept column implies $v_0 = 0$. Hence the matrix is positive definite. Strict convexity gives uniqueness, and differentiation gives the displayed normal equation. $\square$ $\square$

Remark 2.2 The argument establishes computational uniqueness but does not supply observations. In particular, the two-quarter shock chart in the application is regularised rather than statistically identified.

3 Median registration

Proposition 3.1 (Coordinatewise median) *For observations $x_1, \dots, x_N \in \mathbb{R}^p$, the minimisers of*

$$L(a) = \sum_{q=1}^N \|x_q - a\|_1$$

are precisely the Cartesian product of the univariate median intervals of the coordinates. The minimiser is unique if every coordinate has a unique sample median.

Proof Because the ℓ^1 norm is separable,

$$L(a) = \sum_{k=1}^p \sum_{q=1}^N |x_{qk} - a_k|.$$

For a univariate sample, write $n_{<}(u)$, $n_{>}(u)$, and $n_{=}(u)$ for the numbers of observations below, above, and equal to u . The subdifferential of $L_k(u) = \sum_q |x_{qk} - u|$ is

$$\partial L_k(u) = n_{<}(u) - n_{>}(u) + [-n_{=}(u), n_{=}(u)].$$

It contains zero exactly on the median interval. The Cartesian-product statement follows, as does uniqueness. $\square$ $\square$

Example 3.2 (Pure level conversion) Let $x_q = \delta + s_q$ with $\text{med}_q s_q = 0$. The registered translation is δ , however large $\|\delta\|$ may be, and the translated defects are s_q . A large translation is therefore not, by itself, evidence of residual incompatibility.

4 Projection onto the correction set

Theorem 4.1 (Explicit correction) *Let $C \subset \mathbb{R}^p$ be non-empty, closed, and convex. The unique minimiser of*

$$J(c) = \frac{1}{N} \sum_{q=1}^N \|r(q) - c\|_2^2, \quad c \in C,$$

is $\hat{c} = \Pi_C(\bar{r})$, where $\bar{r} = N^{-1} \sum_q r(q)$. If $C = \prod_k [-b_k, b_k]$, then $\hat{c}_k = \min\{b_k, \max\{-b_k, \bar{r}_k\}\}$.

Proof The variance identity gives

$$\frac{1}{N} \sum_q \|r(q) - c\|_2^2 = \frac{1}{N} \sum_q \|r(q) - \bar{r}\|_2^2 + \|\bar{r} - c\|_2^2.$$

The first term is independent of c . The projection theorem for a non-empty closed convex subset of Euclidean space (Boyd and Vandenberghe 2004; Rockafellar and Wets 1998) yields the unique minimiser $\Pi_C(\bar{r})$. Projection onto a Cartesian product of intervals is coordinatewise clipping. $\square$ $\square$

Proposition 4.2 (Variational characterisation and non-expansiveness) *For $x \in \mathbb{R}^p$ and $c = \Pi_C(x)$,*

$$\langle x - c, z - c \rangle \leq 0 \quad \text{for all } z \in C.$$

For all $x, y \in \mathbb{R}^p$,

$$\|\Pi_C(x) - \Pi_C(y)\|_2 \leq \|x - y\|_2.$$

Proof The first statement is the first-order optimality condition for minimising the strictly convex function $z \mapsto \frac{1}{2} \|x - z\|_2^2$ over C . Apply it once with $(x, c, z) = (x, \Pi_C x, \Pi_C y)$ and once with $(y, \Pi_C y, \Pi_C x)$. Adding the inequalities gives

$$\|\Pi_C x - \Pi_C y\|^2 \leq \langle x - y, \Pi_C x - \Pi_C y \rangle,$$

and Cauchy–Schwarz yields the result. $\square$ $\square$

5 Weighted graph decomposition

Theorem 5.1 (Weighted exact-cycle decomposition) *For every $a \in \mathbb{R}^{m \times p}$ there is a unique pair $(\hat{g}, h)$ with $\hat{g}^\top \mathbf{1} = 0$ such that*

$$a = B\hat{g} + h, \quad B^\top W h = 0.$$

Moreover, $B\hat{g}$ and h are orthogonal under $\langle \cdot, \cdot \rangle_W$, and the following are equivalent: (i) $h = 0$; (ii) $a = Bg$ for some g ; (iii) the signed sum of a_e on every cycle is zero.

Proof Consider

$$\min_{g: \mathbf{1}^\top g = 0} \left\| W^{1/2}(a - Bg) \right\|_F^2.$$

On $\mathbf{1}^\perp$, the Hessian $2B^\top WB \otimes I_p$ is positive definite because G is connected. Hence the minimiser $\hat{g}$ is unique. The normal equation is

$$B^\top W(a - B\hat{g}) = 0.$$

Set $h = a - B\hat{g}$. This gives the decomposition and orthogonality to every element of $\text{im } B$; it is the finite weighted analogue of an exact-cycle decomposition on a cellular complex (Bredon 1997; Robinson 2014).

If $h = 0$, then $a = B\hat{g}$, proving (i) $\Rightarrow$ (ii). If $a = Bg$, the signed sum around a cycle telescopes, so (ii) $\Rightarrow$ (iii). To prove (iii) $\Rightarrow$ (ii), fix a root vertex v_0 . For each vertex v , choose a path from v_0 to v and define g_v as the signed sum of a_e along that path. The cycle condition makes this value path-independent. For every oriented edge (i, j) , extending a path to i by that edge shows $a_{ij} = g_j - g_i$. Subtracting the mean of the g_i enforces $\mathbf{1}^\top g = 0$ without changing Bg . $\square$

Corollary 5.2 (Trees) *If G is a tree, every edge translation is exact.*

Proof A tree has no cycles. Condition (iii) is vacuous. $\square$

Example 5.3 (Inconsistent three-agency registry) Let the graph be the oriented triangle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ and let the scalar translations be $a_{12} = 1$, $a_{23} = 1$, and $a_{31} = -1$. Every individual conversion can be applied, but the cycle sum is 1. No offsets g_i satisfy all three equations. The registry changes a value by one unit after a complete circuit.

Example 5.4 (Why a chain still needs residuals) On the chain $1 \rightarrow 2 \rightarrow 3$, every pair of translations is exact, regardless of the size of the observed shape residuals. The vanishing cycle obstruction does not establish empirical comparability. It states only that the declared constant conversions are mutually representable by vertex offsets.

6 Perturbation bounds

Proposition 6.1 (Distance perturbation) *For non-empty compact sets $C, \tilde{C}$ in a Hilbert space and points $x, \tilde{x}$,*

$$|\text{dist}(x, C) - \text{dist}(\tilde{x}, \tilde{C})| \leq \|x - \tilde{x}\| + d_H(C, \tilde{C}).$$

Proof For any $c \in C$, choose $\tilde{c} \in \tilde{C}$ with $\|c - \tilde{c}\| \leq d_H(C, \tilde{C}) + \eta$. Then

$$\text{dist}(\tilde{x}, \tilde{C}) \leq \|\tilde{x} - \tilde{c}\| \leq \|\tilde{x} - x\| + \|x - c\| + d_H(C, \tilde{C}) + \eta.$$

Take the infimum over c , let $\eta \downarrow 0$, and interchange the pairs to obtain the absolute-value bound. $\square$

Proposition 6.2 (Cycle-projection stability) *Let P_{cyc} be the W -orthogonal projection onto $\ker(B^\top W)$. Then*

$$\|P_{\text{cyc}}a - P_{\text{cyc}}\tilde{a}\|_W \leq \|a - \tilde{a}\|_W.$$

Proof Orthogonal projections in a Hilbert space are non-expansive. $\square$ $\square$

7 Full-record lift and dual certificate

Theorem 7.1 (Farkas certificate for the stacked lift) *Exactly one of the following systems has a solution:*

$$Ax \leq b, \quad Ex = h, \quad (P)$$

$$\lambda \geq 0, \quad A^\top \lambda + E^\top y = 0, \quad \lambda^\top b + y^\top h < 0. \quad (D)$$

Consequently, the audited vectors admit a full-record lift if and only if no dual pair (λ, y) satisfying (D) exists.

Proof If x solves (P) and (λ, y) solves the first two conditions of (D), then

$$0 = x^\top (A^\top \lambda + E^\top y) = \lambda^\top Ax + y^\top Ex \leq \lambda^\top b + y^\top h.$$

Thus the strict final inequality in (D) is impossible. Conversely, apply the standard Farkas alternative to the polyhedron obtained by replacing $Ex = h$ with the pair $Ex \leq h$ and $-Ex \leq -h$. Combining the two non-negative multipliers of these inequalities into a free vector y gives (D). $\square$ $\square$

Example 7.2 (Exact audited totals with no admissible lift) Two programmes report audited totals $z_1 = z_2 = 10$. The first full record is $x_1 = (u_1, v_1)$ with $u_1 + v_1 = 10$, $u_1 \leq 4$, and $v_1 \leq 4$. The audit projection is $P_1 x_1 = u_1 + v_1$. The audited value is numerically legitimate but the capacity constraints imply $u_1 + v_1 \leq 8$. Hence no full record lifts $z_1 = 10$. A dual certificate is obtained by adding the two capacity inequalities and subtracting the audited equality.

Example 7.3 (Positive lift) Replace the capacities by $u_1 \leq 6$ and $v_1 \leq 6$. Then $x_1 = (5, 5)$ is a lift of the audited total 10. If an adjoining record has the same audited value and its overlap relation requires equality of the two components, choosing $x_2 = (5, 5)$ supplies a compatible full-record pair.

8 Decision margin theorem

Order the actions as conditional pooling < review < keep separate.

Theorem 8.1 (Margin stability and monotonicity) *Suppose the metadata flag is zero and $|\tilde{D} - D| \leq \varepsilon_D$, $|\tilde{p} - p| \leq \varepsilon_p$. If*

$$\min\{|D - T_L|, |D - T_H|\} > \varepsilon_D, \quad \min\{|p - P_L|, |p - P_H|\} > \varepsilon_p,$$

then the three-way action is unchanged. Increasing any of D , p , or the metadata flag cannot produce a less cautious action.

Proof The margin conditions ensure that neither D nor p crosses any boundary of the rectangular regions that define the rule. Hence the logical truth values of $D > T_H$, $D \geq T_L$, $p > P_H$, and $p \geq P_L$ are unchanged. Monotonicity follows directly from the use of upper sets in these inequalities and from the fact that a metadata flag of one forces the maximal action. $\square$ $\square$

9 Level–shape separation

Proposition 9.1 (Exact linear formula) *Let $t_1, \dots, t_N$ have sample median and mean zero. Suppose*

$$f_i(t) = \alpha + \beta t, \quad f_j(t) = \alpha + \delta + (\beta + \kappa)t.$$

With no correction set, median registration gives $a = \delta$ and

$$D = \frac{\|\kappa\|_2}{\sqrt{p}} \left(\frac{1}{N} \sum_{q=1}^N t_q^2 \right)^{1/2}.$$

Proof The defect is $\Delta(t_q) = \delta + \kappa t_q$. Coordinatewise median zero of t_q gives median δ . After registration, $r(t_q) = \kappa t_q$. Therefore

$$D^2 = \frac{1}{pN} \sum_q \|\kappa t_q\|_2^2 = \frac{\|\kappa\|_2^2}{p} \frac{1}{N} \sum_q t_q^2.$$

Take square roots. $\square$ $\square$

10 Consistency extension

The empirical application uses fixed short windows, so the following asymptotic statement is not invoked to justify its sensitivity bands. It records the population target of the edge procedure in a setting with increasing window length.

Theorem 10.1 (Consistency under stationary ergodic defects) *Let $\{\Delta_q\}_{q \geq 1}$ be a stationary ergodic sequence in $\mathbb{R}^p$ with $E \|\Delta_1\|^2 < \infty$. Assume each coordinate distribution has a unique median a_k^* . Let $\hat{a}_N$ be the coordinatewise sample median, let $C \subset \mathbb{R}^p$ be non-empty compact convex, and define*

$$\hat{c}_N = \Pi_C \left(N^{-1} \sum_{q=1}^N (\Delta_q - \hat{a}_N) \right),$$

$$D_N^2 = \frac{1}{pN} \sum_{q=1}^N \|\Delta_q - \hat{a}_N - \hat{c}_N\|^2.$$

Then $\hat{a}_N \rightarrow a^$ almost surely, $\hat{c}_N \rightarrow c^*$ almost surely, and $D_N \rightarrow D^*$ almost surely, where*

$$c^* = \Pi_C(E[\Delta_1 - a^*]), \quad (D^*)^2 = p^{-1} E \|\Delta_1 - a^* - c^*\|^2.$$

Table 1 Median post-registration residual and share exceeding the illustrative value 0.12

Noise σ	Scenario	Median residual	Share > 0.12
0.02	no break	0.026	0.000
0.02	level only	0.026	0.000
0.02	shape only	0.105	0.001
0.02	level and shape	0.105	0.004
0.05	no break	0.065	0.000
0.05	level only	0.065	0.000
0.05	shape only	0.120	0.501
0.05	level and shape	0.121	0.526
0.10	no break	0.131	0.708
0.10	level only	0.132	0.710
0.10	shape only	0.166	0.981
0.10	level and shape	0.165	0.972

Proof The ergodic theorem applied to indicators $1\{\Delta_{qk} \leq x\}$ at continuity points yields pointwise convergence of the empirical distribution functions to the marginal distributions. Uniqueness of the median implies strong consistency of each sample median, hence $\hat{a}_N \rightarrow a^*$. The vector ergodic theorem and the preceding convergence give

$$N^{-1} \sum_q (\Delta_q - \hat{a}_N) \rightarrow E\Delta_1 - a^*.$$

Continuity and non-expansiveness of Π_C yield $\hat{c}_N \rightarrow c^*$. Expanding the squared norm and using the ergodic theorem for $\|\Delta_q\|^2$, together with convergence of the constant vectors $\hat{a}_N + \hat{c}_N$, gives $D_N^2 \rightarrow (D^*)^2$. Non-negativity gives convergence of the square roots. $\square \square$

11 Controlled simulation details

The simulation uses the grid $t_q \in [-1, 1]$ with eight equally spaced points and three coordinates. The common intercept and slope are

$$\alpha = (0.35, 0.25, 0.20), \quad \beta = (0.08, -0.04, 0.05).$$

The level vector is $\delta = (0.30, -0.20, 0.25)$ and the shape vector is $\kappa = (0.18, -0.12, 0.16)$. The left and right records are

$$Y_{Lq} = \alpha + \beta t_q + \varepsilon_{Lq}, \quad Y_{Rq} = \alpha + \delta_s + (\beta + \kappa_s) t_q + \varepsilon_{Rq},$$

where (δ_s, κ_s) takes the values $(0, 0)$, $(\delta, 0)$, $(0, \kappa)$, and (δ, κ) . The errors are independent $N(0, \sigma^2 I_3)$ with $\sigma \in \{0.02, 0.05, 0.10\}$. Each cell has 1,000 replications. The translation is componentwise median and the correction fraction is 0.15.

At each noise level, the level-only and no-break cases have nearly identical post-registration residuals although their raw discrepancies differ sharply. A fixed tolerance also depends on the noise environment. The empirical thresholds are therefore an illustrative governance registry, not a test of universal size.

12 Reproducibility map

Online Resource 2 contains the exact source-aligned panel used in the analysis, its six-column computational projection, registered parameters, scripts, tables, figures, and checksums. The command

```
python scripts/verify_reported_results.py
```

verifies the source projection and the article’s baseline values. The command

```
python scripts/run_qaq_extension.py
```

regenerates the common-window benchmark diagnostics, the controlled simulation, and the additional figures used in the revised article.

References

- Boyd S, Vandenberghe L (2004) Convex Optimization. Cambridge University Press, Cambridge, <https://doi.org/10.1017/CBO9780511804441>
- Bredon GE (1997) Sheaf Theory, 2nd edn. Springer, New York, <https://doi.org/10.1007/978-1-4612-0647-7>
- Robinson M (2014) Topological Signal Processing. Springer, Berlin, <https://doi.org/10.1007/978-3-662-44104-6>
- Rockafellar RT, Wets RJB (1998) Variational Analysis. Springer, Berlin, <https://doi.org/10.1007/978-3-642-02431-3>