Score Residuals and Tighter Drift Bounds

J. R. Landers

Even with fixed model parameters, predictions can change as inputs drift. How much the score changes depends on how steep the predictor is in the directions the inputs actually move. Drift-aligned tangent regularization (DTR) penalizes sensitivity in a fixed subspace estimated from historical drift [1]. Forecast-weighted DTR replaces that fixed weighting with a conditional velocity second moment [2]. This note asks whether a penalty on the difference between a score's time derivative and the scaled score can sharpen the bound on risk volatility. The answer depends on the residual, the average squared score, and how the score changes between the endpoints.

1. The quantity DTR controls

Let $X_t \in \mathbb{R}^d$, $0 \le t \le T$, be a d-dimensional input process with almost surely absolutely continuous paths over a horizon $T > 0$. Let f be a fixed, continuously differentiable scalar predictor. For a fixed reference score c and an integrable process H_t, write

$$J_t = \nabla f(X_t), \qquad s_t = f(X_t) - c, \qquad \langle H_t \rangle = \frac{1}{T}\int_0^T \mathbb{E}[H_t]\, dt,$$

with expectation conditional on predeployment fitting. Dots denote time derivatives. By the chain rule, the score's rate of change is the dot product of input velocity and the score gradient. Assume s_0 is square integrable and the energy $\mathcal{A}$ is finite.

$$\dot{s}_t = J_t^\top \dot{X}_t, \qquad \mathcal{A} := \langle \dot{s}_t^2 \rangle < \infty.$$

Since $s_t = s_0 + \int_0^t \dot{s}_u\, du$, Cauchy–Schwarz also gives square integrability of s_T and finiteness of $\langle s_t^2 \rangle$.

Let $g(x)$ be the conditional expected loss at input x and $r(t) = \mathbb{E}[g(X_t)]$. Assume the along-path regularity conditions of [1] and the directional bound

$$|\nabla g(X_t)^\top \dot{X}_t| \le \beta |\dot{s}_t|,$$

where $\beta \ge 0$ bounds how strongly score changes translate into loss changes along the path. Jensen's inequality then gives $|r'(t)|^2 \le \beta^2 \mathbb{E}[\dot{s}_t^2]$. The interval Poincaré inequality turns this control of the rate of change into a bound on how much risk varies over time. With U uniform on $[0, T]$, it gives

$$\mathrm{Var}(r(U)) \le \frac{\beta^2 T^2}{\pi^2}\, \mathcal{A}. \tag{1}$$

For the fixed-subspace bound, let the columns of $V \in \mathbb{R}^{d\times k}$ form an orthonormal basis for the k-dimensional drift subspace. The velocity splits as

$$\dot{X}_t = V a_t + \rho_t, \qquad a_t = V^\top \dot{X}_t, \qquad \rho_t = P_\perp \dot{X}_t, \qquad P_\perp = I - VV^\top.$$

Here a_t gives the aligned velocity coordinates and ρ_t is the orthogonal velocity. Let A_0, R_0 bound their squared Euclidean norms, respectively, almost surely for almost every t. Corollary 2 of [1] gives

$$\mathcal{A} \le B_{\mathrm{DTR}} := 2\big(A_0 \left\langle \left\| V^\top J_t \right\|^2 \right\rangle + R_0 \left\langle \|P_\perp J_t\|^2 \right\rangle\big). \tag{2}$$

The factor of two comes from bounding the square of the sum of the aligned and orthogonal contributions. It drops out when $\rho_t = 0$.

2. The score residual and its endpoint correction

For $\alpha > 0$ with units of inverse time, define the squared residual, score energy, and endpoint change by

$$\mathcal{R}_\alpha = \left\langle (\dot{s}_t - \alpha s_t)^2 \right\rangle, \qquad S = \left\langle s_t^2 \right\rangle, \qquad \Delta = \mathbb{E}[s_T^2 - s_0^2].$$

The residual is a squared L^2 discrepancy of the kind used in differential-equation smoothing [3]. Expanding the square and integrating $2s_t\dot{s}_t = (s_t^2)'$ pathwise gives

$$\mathcal{A} = \mathcal{R}_\alpha - \alpha^2 S + \frac{\alpha}{T}\Delta. \tag{3}$$

The cross term is integrable since $\langle |s_t\dot{s}_t| \rangle \le \sqrt{S\mathcal{A}}$. The endpoint term accounts for net change in the squared score. For $s_t = s_0 e^{\alpha t}$ with $s_0 \neq 0$, the residual is zero because the score follows the target law exactly, while $\mathcal{A} = \alpha^2 S > 0$.

3. A forecast-calibrated certificate

Let $b(t,x)$ and $\Sigma(t,x)$ be the conditional mean and covariance of $\dot{X}_t$ given $X_t = x$, and let $\hat{b}$ and $\hat{\Sigma} \succeq 0$ be their forecasts. All four are evaluated at (t, X_t) below. The forecast-weighted note uses the forecast second moment $\hat{M}$ and its penalty $\mathcal{R}_{\hat{M}}$,

$$\hat{M} = \hat{b}\hat{b}^\top + \hat{\Sigma}, \qquad \mathcal{R}_{\hat{M}} = \left\langle J_t^\top \hat{M} J_t \right\rangle.$$

Conditioning on X_t and applying the variance decomposition instead gives

$$\mathcal{R}_\alpha = \left\langle (J_t^\top b - \alpha s_t)^2 \right\rangle + \left\langle J_t^\top \Sigma J_t \right\rangle. \tag{4}$$

The first term measures the mismatch in mean score velocity. The second measures spread around that mean. Reversing a deterministic velocity leaves its second moment intact but flips its cross term with the score.

Suppose calibration yields

$$\left\langle [J_t^\top (b - \hat{b})]^2 \right\rangle \le \eta^2, \qquad \left\langle J_t^\top \Sigma J_t \right\rangle \le C_\Sigma,$$

where $\eta \ge 0$ bounds the L^2 error in the forecast conditional mean of $\dot{s}_t$, and C_Σ bounds its average conditional variance. Uniform mean-error and covariance bounds on $\dot{X}_t$ give these bounds after projection onto J_t. Set $\hat{D}_\alpha = \left\langle (J_t^\top \hat{b} - \alpha s_t)^2 \right\rangle$. The L^2 triangle inequality in (4) gives

$$\mathcal{R}_\alpha \le (\sqrt{\hat{D}_\alpha} + \eta)^2 + C_\Sigma.$$

Score energy is subtracted and endpoint change is added in (3). Thus deployment bounds $S \ge S_-$ and $\Delta \le \Delta_+$ give

$$\mathcal{A} \le B_{\mathrm{res}} := (\sqrt{\hat{D}_\alpha} + \eta)^2 + C_\Sigma - \alpha^2 S_- + \frac{\alpha}{T}\Delta_+. \tag{5}$$

The bound from [2] is

$$B_{\mathrm{fw}} = \kappa \mathcal{R}_{\hat{M}} + \varepsilon^2 \left\langle \|J_t\|^2 \right\rangle \quad \text{when} \quad \mathbb{E}[\dot{X}_t \dot{X}_t^\top \mid X_t] \preceq \kappa \hat{M} + \varepsilon^2 I$$

almost surely at almost every t, where $\preceq$ denotes positive semidefinite order, $\kappa \ge 1$, and $\varepsilon \ge 0$. In (1), replace $\mathcal{A}$ with the minimum of the available bounds $B_{\mathrm{DTR}}, B_{\mathrm{fw}}, B_{\mathrm{res}}$. For $\beta > 0$, the certificate is strictly tighter whenever B_{res} lies below the previous minimum. Exact trajectories already determine $\mathcal{A}$ through (3). The gain comes when the residual gives a better upper bound on that same energy.

4. A seasonal example and a training objective

Take $T = 1$, $X_t = \sin(2\pi t)$, $f(x) = x$, and $c = 0$. The one-dimensional drift direction is fixed, and the peak squared speed gives $B_{\mathrm{DTR}} = 4\pi^2$ with $\rho_t = 0$. With exact forecasts,

$$\mathcal{R}_\alpha = 2\pi^2 + \frac{\alpha^2}{2}, \qquad S = \frac{1}{2}, \qquad \Delta = 0.$$

The seasonal path returns to its starting score, so the endpoint correction vanishes. The corrected bound is $2\pi^2$, half the peak-speed bound and equal to the exact forecast-weighted energy. The gap comes from using the average squared speed over the cycle instead of its peak. With a mean forecast error allowance η around that forecast and a covariance allowance C_Σ, it still beats the peak-speed bound whenever

$$2\eta\sqrt{2\pi^2 + \alpha^2/2} + \eta^2 + C_\Sigma < 2\pi^2,$$

assuming exact S, Δ.

A training objective can add the term

$$\mu\big(\hat{D}_\alpha + \left\langle J_t^\top \hat{\Sigma} J_t \right\rangle\big)$$

to supervised loss and the DTR penalty, with penalty weight $\mu \geq 0$. The two penalties encode different goals. The residual favors the score law $\dot{s}_t = \alpha s_t$, while DTR favors predictions that change little along nuisance motion. A deployment certificate requires bounds on the forecast and trajectory averages in (5), including sampling error when those averages are estimated. The certificate controls temporal risk variance. Average loss remains a separate quantity to evaluate.

References

[1] J. R. Landers, *Jacobian–Velocity Bounds for Deployment Risk Under Covariate Drift*, 2026. Repository.

[2] J. R. Landers, *Forecast-Weighted Drift Regularization*, 2026. Companion note.

[3] J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao, Parameter estimation for differential equations: a generalized smoothing approach, *J. R. Stat. Soc. B*, 69(5), 741–796, 2007. DOI.