Summary — Online Kernel Supervised Dimensionality Reduction

Problem. Kernel-based supervised DR is powerful but typically impractical online due to quadratic Gram costs; linear SDR misses nonlinear dependencies common in real data.

Method. I propose OKSDR, a single-pass, fixed-memory algorithm that learns a task-aligned nonlinear subspace U by maximizing centered input–output cross-covariance. The method uses explicit random Fourier features for $\phi(x)$ and optionally $\psi(y)$, maintains running means and a raw cross-moment to form C_{XY}, and performs per-sample orthogonalized Stiefel/QR updates with optional Adam-style clipping. No replay buffers or Gram matrices are stored.

Objective & Update. Maximize $f(U) = |C_{XY}^{\top} U|_F^2$; compute Euclidean $\nabla_E f$, project to tangent $\nabla_R f$, take a first-order step, then QR-retract to preserve $U^{\top} U = I$.

Complexity & Memory. Per sample time is dominated by RFF transforms, C_{XY} update, gradient multiplication, and occasional QR; stored state scales as $O(D_x D_y) + O(D_x k)$, independent of stream length n.

Experiments. On eight synthetic datasets and Kin8nm, OKSDR retains 89–99% (median ≈95%) of the batch version, is typically more noise-robust than linear baselines. On Kin8nm, strict nested CV shows online nearly matches batch while exceeding linear/shallow baselines by ~0.15 absolute R^2.

Adaptation. Supports unbiased averaging $(\gamma_t = 1/t)$ or exponential forgetting $(\gamma_t = 1 - \lambda)$ to track non-stationarity (effective window $\approx 1/(1 - \lambda)$).

Takeaway. OKSDR delivers nonlinear, supervised representations in streaming settings with single-pass updates and fixed memory, reaching near-batch accuracy at costs comparable to linear models.