PREDICTION TECHNIQUES FOR DYNAMIC IMAGING WITH ONLINE PRIMAL–DUAL METHODS

Neil Dizon*   Jyrki Jauhiainen†   Tuomo Valkonen‡

Abstract   Online optimisation facilitates the solution of dynamic inverse problems, such as image stabilisation, fluid flow monitoring, and dynamic medical imaging. In this paper, we improve upon previous work on predictive online primal-dual methods on two fronts. Firstly, we provide a more concise analysis that symmetrises previously unsymmetric regret bounds, and relaxes previous restrictive conditions on the dual predictor. Secondly, based on the latter, we develop several improved dual predictors. We numerically demonstrate their efficacy in image stabilisation and dynamic positron emission tomography.

1 INTRODUCTION

Many real-world applications involve processing information evolving over time. This includes tasks like computational image stabilisation based on rapid successions of noisy images [30, 33, 38], fluid flow monitoring in industrial processes [3, 15, 16, 20], as well as the reconstruction of medical images in the presence of physical motion [5, 6, 17, 21]. When the monitoring period is long, and the results are needed immediately, while data is still arriving, it is not feasible to solve one large reconstruction problem after all the data has arrived. Instead, online reconstruction techniques are required.

Online optimisation extends traditional optimisation by allowing the objective function, parameters, or constraints to change over time, with each iteration of the algorithm. In this paper, we consider the formal problem

$$(1.1) \qquad \min_{(x^0,x^1,x^2\dots)\in\mathcal{X}} \sum_{k=0}^{\infty} J_k := F_k(x^k) + E_k(x^k) + G_k(K_k x^k),$$

where $F_k, E_k : X_k \to \overline{\mathbb{R}}$, $G_k : Y_k \to \overline{\mathbb{R}}$ are convex, proper, and lower semi-continuous on Hilbert spaces X_k and Y_k ($k \in \mathbb{N}$), E_k is additionally smooth, and $K_k \in \mathbb{L}(X_k; Y_k)$ is linear and bounded. The set $\mathcal{X} \subset \prod_{k=0}^{\infty} X_k$ encodes temporal coupling between the variables, and the frame index k represents time-evolution. For example, basing computational image stabilisation on consecutive temporally coupled total variation denoising problems, we arrive at

$$\min_{(x^0,x^1,x^2\dots)\in\mathcal{X}} \sum_{k=0}^{\infty} \frac{1}{2}\|x^k - z_k\|^2 + \alpha\|D_k x^k\|_{2,1},$$

*Department of Mathematics and Statistics, University of Helsinki, Finland, neil.dizon@helsinki.fi, ORCID: 0000-0001-8664-2255

†Department of Technical Physics, University of Eastern Finland, jyrki.h.jauhiainen@helsinki.fi, ORCID: 0000-0001-6711-6997

‡ModeMat, Escuela Politécnica Nacional, Quito, Ecuador and Department of Mathematics and Statistics, University of Helsinki, Finland, tuomo.valkonen@iki.fi, ORCID: 0000-0001-6683-3572

where z_k is the measurement data, and α is a regularisation parameter for isotropic total variation based on the (discretised) differential operator D_k. The set $\mathcal{X}$ models, for example, the optical flow between consecutive image frames. Similarly, dynamic positron emission tomography (PET) reconstruction affected by patient body motion can be modelled as

$$\min_{(x^0, x^1, x^2 \ldots) \in \mathcal{X}} \sum_{k=0}^{\infty} \delta_{\geq 0}(x^k) + \sum_{i=1}^{n} \left([A_k x^k]_i - [z_k]_i \log([A_k x^k + c_k]_i) \right) + \alpha \|D_k x^k\|_{2,1},$$

where A_k is the forward model based on a partial Radon transform and c_k is a known vector with non-negative entries.

Early dynamic (consensus-type) optimisation studied series of static problems [19, 23], assuming sufficient computational resources to solve for each k the individual static problems

$$(1.2) \qquad \min_{x^k \in X_k} F_k(x^k) + E_k(x^k) + G_k(K_k x^k),$$

before new data arrives. However, such an approach fails to exploit for temporal super-resolution any physical temporal coupling present between the data for the different frames k. Alternatively, it is possible to solve for each increasing N the finite time window problem

$$\min_{x^{0:N} \in \mathcal{X}_{0:N}} \sum_{k=0}^{N} F_k(x^k) + E_k(x^k) + G_k(K_k x^k).$$

where we use the shorthand slicing notations $x^{0:N} := (x^0, \ldots, x^N)$ and $\mathcal{X}_{0:N} := \{x^{0:N} \mid (x^0, x^1, \ldots) \in \mathcal{X}\}$. However, for large N, these problems become numerically prohibitively expensive. Memory constraints may also compel the exclusion of earlier data, thereby limiting the approach to short windows of recent data. We will thus adopt an online optimisation approach, focusing on cases where one can only afford one (or at most a few) steps of an optimisation algorithm within a time interval of interest. For modern introductions to online optimisation, we refer to [1, 13, 24].

Online optimisation algorithms for dynamic problems are categorised into *structured* and *unstructured* methods [26]. Structured algorithms take advantage of the temporal nature of the problem to predict and approximate an optimal solution at each time step. On the other hand, unstructured algorithms are agnostic of the temporal nature of the problem and rely only on the optimisation problem that are presented at each time, or a history thereof. Non-predictive primal-dual methods under this category include [4, 29, 36]. Structured online algorithms can be further subcategorised into *prediction-correction* methods and *predictors*. A mere predictor carries out one step of an optimisation algorithm with respect to a predicted objective function but does not perform a corrective step when the new problem becomes available, e.g., see [9, 22, 35]. In contrast, prediction-correction methods predict how the optimisation problem changes, and then correct for the errors in predictions once the new objective function is revealed. Primal methods under this category include [12, 27, 28, 37] based on gradient and mirror descents, among others. The available literature on primal-dual methods within this class is limited to [25, 33]. In particular, the Predictive Online Primal-Dual Proximal Splitting (POPD) method of [33] is amenable to the online solution of (1.1) where $E_k = 0$.

The difficulty with proving something about online methods is that *convergence* results are rarely available. Instead, one attempts to bound the *regret* of past updates with respect to all information available up to an instant N. In the dynamic case, following [12], one bounds the *dynamic regret* defined by

$$\text{dynamic\_regret}(x^{0:N}) = \sup_{\bar{x}^{0:N} \in \mathcal{X}_{0:N}} \sum_{k=0}^{N} \left(J_k(x^k) - J_k(\bar{x}^k) \right).$$

This regret may be negative, if the *comparison set* $\mathcal{X}_{0:N}$ constrains the *comparison sequence* $\bar{x}^{0:N}$ more than the predictions and updates constrain the iterates. Alternatively, performance evaluation based on *asymptotical tracking errors* is available in the literature but its discussion is not included here. For details, interested reader is referred to, e.g., [26]. For the POPD of [33], the regret estimate is even weaker: for a function $\check{J}_{0:N}$ dependent on both the comparison set $\mathcal{X}_{0:N}$ and the original objectives $J_0, \ldots, J_N$, we only have a bound on

$$(1.3) \qquad \sup_{\bar{x}^{0:N} \in \mathcal{X}_{0:N}} \left(\check{J}_{0:N}(x^{0:N}) - \sum_{k=0}^{N} J_k(\bar{x}^k) \right).$$

That is, the regret estimate is non-symmetric between the algorithmic iterates and the comparison sequences.

Due to proof-technical reasons, the dual predictor of the POPD in [33] is also severely constrained to a specific proximal form. In Section 2 of this paper, through improved and much simplified proofs, we (a) remove this restrictions, (b) provide improved, symmetric, regret estimates and, (c) extend the POPD with an additional forward step with respect to E_k (which was zero in [33]). Specifically, our new form of dynamic regret for primal-dual methods symmetrises (1.3) by replacing both $\check{J}_k$ and J_k with a temporal "sub-infimal" convolution $\mathring{J}_k$ between the dual comparison set and the objective. The bound for this modified dynamic regret depends on the richness of the comparison set and the accuracy of the predictors. Given the relaxed conditions on the dual predictor, in Section 3, we analyse the accuracy of a broad class of "pseudo-affine" primal-dual predictors. We then present various examples of pseudo-affine predictors in the context of optical flow, and how they preserve salient relationships between the primal and dual variables. Finally in Section 4, we evaluate the proposed method and predictors numerically on image stabilisation and dynamic PET reconstruction.

NOTATION

We write $x^{n:m} := (x^n, \ldots, x^m)$ with $n \le m$, and $x^{n:\infty} := (x^n, x^{n+1}, \ldots)$. We slice a set $\mathcal{X} \subset \prod_{k=0}^{\infty} X_k$ as $\mathcal{X}_{n:m} := \{x^{n:m} \mid x^{0:\infty} \in \mathcal{X}\}$ and $\mathcal{X}_n := \mathcal{X}_{n:n}$. We write $\mathbb{L}(X; Y)$ for the set of bounded linear operators between (Hilbert) spaces X and Y, and $\mathrm{Id} \in \mathbb{L}(X; X)$ for the identity operator. For brevity, we write $\langle x, y \rangle_M := \langle Mx, y \rangle$, and $[\![x]\!]_M := \sqrt{\langle x, x \rangle}_M$ for $M \in \mathbb{L}(X; X)$. We write $M \ge 0$ if M is positive semi-definite and $M \simeq N$ if $\langle Mx, x \rangle = \langle Nx, x \rangle$ for all x. When M is positive semi-definite, we use the usual norm notation $\|x\|_M := \sqrt{\langle x, x \rangle_M}$.

For any $A \subset X$ and $x \in X$ we set $\langle A, x \rangle := \{ \langle z, x \rangle \mid z \in A \}$. We write δ_A for the $\{0, \infty\}$-valued indicator function of A. For any $B \subset \mathbb{R}$ (in particular $B = \langle A, x \rangle$), we use the notation $B \ge 0$ to mean that $t \ge 0$ for all $t \in B$.

For $F : X \to (-\infty, \infty]$, the effective domain $\mathrm{dom}\, F := \{x \in X \mid F(x) < \infty\}$. With $\overline{\mathbb{R}} := [-\infty, \infty]$ the set of extended reals, we call $F : X \to \overline{\mathbb{R}}$ *proper* if $F > -\infty$ and $\mathrm{dom}\, F \ne \emptyset$. Let then F be convex. We write $\partial F(x)$ for the subdifferential at x and (for additionally proper and lower semicontinuous F)

$$\mathrm{prox}_F(x) := \arg\min_{\tilde{x} \in X} F(\tilde{x}) + \frac{1}{2} \|\tilde{x} - x\|^2 = (\mathrm{Id} + \partial F)^{-1}(x)$$

for the proximal map. We call F strongly subdifferentiable at x with the factor $\gamma > 0$ if

$$F(\tilde{x}) - F(x) \ge \langle z, \tilde{x} - x \rangle + \frac{\gamma}{2} \|\tilde{x} - x\|^2 \quad \text{for all} \quad z \in \partial F(x) \text{ and } \tilde{x} \in X.$$

In Hilbert spaces, this is equivalent to strong convexity with the same factor. Finally, for $f \in L^q(\Omega; \mathbb{R}^n)$, we write $\|f\|_{p,q} := \left\| \xi \mapsto \|f(\xi)\|_p \right\|_{L^q(\Omega)}$.

Algorithm 1 New predictive online primal-dual proximal splitting (POPD$_2$)

Require: For all $k \in \mathbb{N}$, on Hilbert spaces X_k and Y_k, convex, proper, lower semi-continuous $F_{k+1}, E_{k+1} :$
$X_{k+1} \to \overline{\mathbb{R}}$ and $G_{k+1}^* : Y_{k+1} \to \overline{\mathbb{R}}$, primal-dual predictors $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$, and
$K_{k+1} \in \mathbb{L}(X_{k+1}; Y_{k+1})$. Step length parameters $\tau_{k+1}, \sigma_{k+1} > 0$.
1: Pick initial iterates $x^0 \in X_0$ and $y^0 \in Y_0$.
2: **for** $k \in \mathbb{N}$ **do**
3:      $(\check{x}^{k+1}, \check{y}^{k+1}) := P_k(x^k, y^k)$. $\rightsquigarrow$ *prediction step*
4:      $x^{k+1} := \operatorname{prox}_{\tau_{k+1}F_{k+1}}(\check{x}^{k+1} - \tau_{k+1}\nabla E_{k+1}(\check{x}^{k+1}) - \tau_{k+1}K_{k+1}^*\check{y}^{k+1})$ $\rightsquigarrow$ *primal step*
5:      $y^{k+1} := \operatorname{prox}_{\sigma_{k+1}G_{k+1}^*}(\check{y}^{k+1} + \sigma_{k+1}K_{k+1}(2x^{k+1} - \check{x}^{k+1}))$ $\rightsquigarrow$ *dual step*
6: **end for**

---

2 AN ONLINE PRIMAL-DUAL METHOD

In this section we present and analyse our proposed POPD$_2$ primal-dual method for the online solution of (1.1). Presented in Algorithm 1, the method incorporates an additional forward step with respect to E_k, and simplifies the dual prediction of the POPD of [33]. Each step k of the algorithm corresponds to a single data frame, with the time-varying data embedded in the functions and operators (F_k, E_k, K_k, G_k, and P_k).[1] The primal and dual steps (Lines 4 and 5) are analogous to the standard PDPS of Chambolle and Pock [8] with an additional forward step with respect to E_k. These optimisation steps are preceded by a prediction step (Line 3), implemented by the operators $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$ that transfer iterates from one step to the next.

In Section 2.1 we outline our assumptions. We then present in Section 2.2 a symmetric dynamic regret bound, as discussed in the introduction.

2.1 ASSUMPTIONS AND DEFINITIONS

To develop the regret theory, we work with the testing approach to convergence proofs, presented in [31, 32, 10]. This depends on encoding convergence rates into distinct testing parameters for the primal and dual variables. With the general notation $u = (x, y)$, $u^k = (x^k, y^k)$, etc., we work with the following assumptions.

Assumption 2.1. For all $k \geq 1$, on Hilbert spaces X_k and Y_k, we are given:

(i) Convex, proper, and lower semicontinuous $F_k, E_k : X_k \to \overline{\mathbb{R}}$, $G_k^* : Y_k \to \overline{\mathbb{R}}$, as well as $K_k \in \mathbb{L}(X_k; Y_k)$, such that ∇E_k exists and is L_k-Lipschitz. We write $Q_k := F_k + E_k$, and $\gamma_{F_k}, \gamma_{E_k}, \rho_k \geq 0$ for the factors of (strong) convexity of F_k, E_k and G_k^*, respectively. For some $\kappa_k \in (0, 1]$ we have

$$(2.1) \qquad 0 \leq \gamma_k := \begin{cases} \gamma_{F_k} + \gamma_{E_k} - \kappa_k L_k, & \gamma_{E_k} > 0, \\ \gamma_{F_k}, & \gamma_{E_k} = 0. \end{cases}$$

(ii) Primal and dual step length parameters $\tau_k, \sigma_k > 0$ and testing parameters $\eta_k, \varphi_k, \psi_k > 0$ satisfying

$$(2.2a) \qquad \eta_k = \varphi_k \tau_k = \psi_k \sigma_k, \qquad \text{(primal-dual coupling)}$$
$$(2.2b) \qquad 1 \geq \tau_k \kappa_k^{-1} L_k + \tau_k \sigma_k \|K_k\|^2 \qquad \text{(metric positivity)}.$$

(iii) Primal-dual predictors $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$ giving the predictions $(\check{x}^{k+1}, \check{y}^{k+1}) := P_k(x^k, y^k)$.

---

[1]It is, of course, possible for each of the functions to depend on multiple real data frames simply by treating them as a single frame in the algorithm, and correspondingly taking the spaces X_k and Y_k larger. They can even grow and shrink with k. These possibilities are exploited [33, Section 5.2].

(iv) A bounded set $\mathcal{U} \subset \prod_{k=0}^{\infty} X_k \times Y_k$ of primal-dual comparison sequences with which we define the set of primal and dual comparison sequences as

$$\mathcal{X} := \left\{ \bar{x}^{0:\infty} \in \prod_{k=0}^{\infty} X_k \mid (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathcal{U} \right\} \quad \text{and}$$
$$\mathcal{Y} := \left\{ \bar{y}^{0:\infty} \in \prod_{k=0}^{\infty} Y_k \mid (\bar{x}^{0:\infty}, \bar{y}^{0:\infty}) \in \mathcal{U} \right\}.$$

Example 2.2. Similarly to the standard PDPS, as analysed in [31, 10], for an unaccelerated method, we can choose the step length and testing parameters as $\tau_k \equiv \tau$ and $\sigma_k \equiv \sigma$ for some $\tau, \sigma > 0$, along with $\eta_k \equiv \tau$, $\varphi_k \equiv 1$ and $\psi_k \equiv \frac{\tau}{\sigma}$. For an accelerated method, more elaborate choices are needed.

For all $k \geq 1$, we define $M_k, \Gamma_k, \Omega_k \in \mathbb{L}(X_k \times Y_k; X_k \times Y_k)$ by

$$M_k := \begin{pmatrix} \tau_k^{-1} \operatorname{Id} & -K_k^* \\ -K_k & \sigma_k^{-1} \operatorname{Id} \end{pmatrix}, \quad \Gamma_k := \eta_k \begin{pmatrix} \gamma_k \operatorname{Id} & 2K_k^* \\ -2K_k & \rho_k \operatorname{Id} \end{pmatrix}, \text{ and } \Omega_k := \begin{bmatrix} \kappa_k^{-1} L_k & 0 \\ 0 & 0. \end{bmatrix}.$$

We further define for all $k \in \mathbb{N}$ the monotone operator $H_k : X_k \times Y_k \rightrightarrows X_k \times Y_k$ as

$$(2.3) \qquad\qquad H_k(u) := \begin{pmatrix} \partial F_k(x) + \nabla E_k(x) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix}.$$

Then $0 \in H_k(\hat{u}^k)$ encodes the first order necessary and sufficient optimality conditions for the static problem (1.2) for frame k. Moreover, writing $u^k := (x^k, y^k)$ and $(\check{x}^{k+1}, \check{y}^{k+1}) := P_k(x^k, y^k)$, Algorithm 1 reads [31, 14] in implicit form

$$(2.4) \qquad\qquad 0 \in \tilde{H}_k(u^k) + M_k(u^k - \check{u}^k) \quad \text{for all} \quad k \geq 1,$$

where $\tilde{H}_k : X_k \times Y_k \rightrightarrows X_k \times Y_k$ is defined by a modification of H_k as

$$(2.5) \qquad\qquad \tilde{H}_k(u) := \begin{pmatrix} \partial F_k(x) + \nabla E_k(\check{x}^k) + K_k^* y \\ \partial G_k^*(y) - K_k x \end{pmatrix}.$$

(Alternatively, to avoid introducing $\tilde{H}_k$, we could replace M_k by a Bregman divergence [32].)

For brevity, with $u^{0:N} = (u^0, \ldots, u^N)$, we also write

$$H_{0:N}(u^{0:N}) := H_0(u^0) \times \cdots \times H_N(u^N), \qquad\qquad G_{1:N}(y^{1:N}) := \sum_{k=1}^{N} \eta_k G_k(\eta_k^{-1} y^k),$$

$$K_{1:N} x^{1:N} := (\eta_1 K_1 x^1, \ldots, \eta_N K_N x^N), \quad \text{and} \qquad Q_{1:N}(x^{1:N}) := \sum_{k=1}^{N} \eta_k [F_k + E_k](x^k).$$

In the setting of Example 2.2, $\eta_k \equiv \tau$, so all the functions are simply scaled by the constant primal step length τ. Observe that $G_{1:N}^*(y^{1:N}) = \sum_{k=1}^{N} \eta_k G_k^*(y^{k+1})$ and

$$[Q_{1:N} + G_{1:N} \circ K_{1:N}](x^{1:N}) = \sum_{k=1}^{N} \eta_k [Q_k + G_k \circ K_k](x^{k+1}).$$

Finally, for each k, we define the *Lagrangian duality gap* by

$$\mathcal{G}_k^H(u^k, \bar{u}^k) := \eta_k \left[F_k(x^k) + E_k(x^k) + \langle K_k x^k, \bar{y}^k \rangle - G_k^*(\bar{y}^k) \right]$$
$$- \eta_k \left[F_k(\bar{x}^k) + E_k(\bar{x}^k) + \langle K_k^* y^k, \bar{x}^k \rangle - G_k^*(y^k) \right].$$

This is non-negative if $0 \in H_k(\bar{u}^k)$; see, e.g., [10].

2.2 A GENERAL REGRET ESTIMATE

As we need to develop a dynamic regret theory for Algorithm 1, we first revisit relevant tools to derive meaningful measures of regret. We first recall the following *smoothness three-point inequalities* (on a Hilbert space X).

Lemma 2.3 ([31, Appendix B] or [10, Chapter 7]). *Suppose* $E : X \to \overline{\mathbb{R}}$ *is convex, proper, and lower semicontinuous, and has* L-*Lipschitz gradient. Then*

$$(2.6) \qquad \langle \nabla E(z), x - \bar{x} \rangle \geq E(x) - E(\bar{x}) - \frac{L}{2}\|x - z\|^2 \quad (\bar{x}, z, x \in X).$$

If E *is, moreover,* γ_E-*strongly convex, then for any* $\beta > 0$ *and* $\bar{x}, z, x \in X$, *also*

$$(2.7) \qquad \langle \nabla E(z), x - \bar{x} \rangle \geq E(x) - E(\bar{x}) + \frac{\gamma_E - \beta L^2}{2}\|x - \bar{x}\|^2 - \frac{1}{2\beta}\|x - z\|^2.$$

Corollary 2.4. *Let Assumption 2.1 hold. Then, for any* $k \in \mathbb{N}$,

$$\langle \partial F_k(x^k) + \nabla E_k(\check{x}^k), x^k - \bar{x}^k \rangle \geq Q_k(x^k) - Q_k(\bar{x}^k) + \frac{\gamma_k}{2}\|x^k - \bar{x}^k\|^2 - \frac{L_k}{2\kappa_k}\|x^k - \check{x}^k\|^2.$$

Proof. If $\gamma_{E_k} = 0$, (2.6) of Lemma 2.3 with the (strong) convexity of F_k yields

$$\langle \partial F_k(x^k) + \nabla E_k(\check{x}^k), x^k - \bar{x}^k \rangle \geq Q_k(x^k) - Q_k(\bar{x}^k) + \frac{\gamma_{F_k}}{2}\|x^k - \bar{x}^k\|^2 - \frac{L_k}{2}\|x^k - \check{x}^k\|^2.$$

If $\gamma_{E_k} > 0$, (2.7) of Lemma 2.3 with $\beta_k = \kappa_k L_k^{-1}$ and the (strong) convexity of F_k yield the estimate

$$\langle \partial F_k(x^k) + \nabla E_k(\check{x}^k), x^k - \bar{x}^k \rangle \geq Q_k(x^k) - Q_k(\bar{x}^k) + \frac{\gamma_{F_k} + \gamma_{E_k} - \kappa_k L_k}{2}\|x^k - \bar{x}^k\|^2 - \frac{L_k}{2\kappa_k}\|x^k - \check{x}^k\|^2.$$

In both cases, the claim follows after an application of (2.1). $\qquad \square$

Using the preceding lemma, the next result bounds the cumulative sum of Lagrangian duality gaps for the iterates of Algorithm 1. This bound is instrumental in proving our main dynamic regret bound, to follow. The topic of the next Section 3 is to bound the prediction error e_N present in the result.

Lemma 2.5. *Let Assumption 2.1 hold for* $u^{1:N}$ *generated by Algorithm 1 for an initial* $u^0 \in X_0 \times Y_0$. *Then* $M_k, \eta_k M_k + \Gamma_k$ *and* $\eta_k(M_k - \Omega_k)$ *are positive semi-definite, and*

$$\frac{1}{2}\|u^N - \bar{u}^N\|^2_{\eta_N M_N + \Gamma_N} + \sum_{k=1}^{N}\left(G_k^H(u^k, \bar{u}^k) + \frac{1}{2}\|u^k - \check{u}^k\|^2_{\eta_k(M_k - \Omega_k)}\right) \leq \frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0 M_0 + \Gamma_0} + e_N(u^{0:N-1}, \bar{u}^{0:N}),$$

for the prediction error

$$(2.8) \qquad e_N(u^{0:N-1}, \bar{u}^{0:N}) := \sum_{k=1}^{N}\left(\frac{1}{2}\|\check{u}^k - \bar{u}^k\|^2_{\eta_k M_k} - \frac{1}{2}\|u^{k-1} - \bar{u}^{k-1}\|^2_{\eta_{k-1}M_{k-1} + \Gamma_{k-1}}\right).$$

Proof. For brevity, and to not abuse norm notation when Γ_k is not positive semi-definite, we write $[\![x]\!]^2_{\Gamma_k} := \langle x, x \rangle_{\Gamma_k}$. By Young's inequality, we have

$$\eta_k M_k - \eta_k \Omega_k = \eta_k \begin{pmatrix} (\tau_k^{-1} - \kappa_k^{-1}L_k)\,\mathrm{Id} & -K_k^* \\ -K_k & \sigma_k^{-1}\mathrm{Id} \end{pmatrix} \geq \varphi_k \begin{pmatrix} \mathrm{Id} - \tau_k\kappa_k^{-1}L_k - \tau_k\sigma_k K_k^*K_k & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\eta_k M_k + \Gamma_k \simeq \begin{pmatrix} \varphi_k(1 + \gamma_k \tau_k) \operatorname{Id} & -\eta_k K_k^* \\ -\eta_k K_k & \psi_k(1 + \rho_k \sigma_k) \operatorname{Id} \end{pmatrix} \geq \begin{pmatrix} \varphi_k(1 + \gamma_k \tau_k) \operatorname{Id} - \frac{\eta_k^2}{\psi_k(1 + \rho_k \sigma_k)} K_k^* K_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus, (2.2b) establishes the positive semi-definiteness claims.

We then expand

$$\eta_k \langle \tilde{H}_k(u^k), u^k - \bar{u}^k \rangle = \eta_k \langle \partial F_k(x^k), x^k - \bar{x}^k \rangle + \eta_k \langle \nabla E_k(\check{x}^k), x^k - \bar{x}^k \rangle$$
$$+ \eta_k \langle \partial G_k^*(y^k), y^k - \bar{y}^k \rangle$$
$$+ \eta_k \langle K_k^* y^k, x^k - \bar{x}^k \rangle - \eta_k \langle K_k x^k, y^k - \bar{y}^k \rangle.$$

Thus the (strong) convexity of F_k and G_k^* together with Corollary 2.4 yields

$$\eta_k \langle \tilde{H}_k(u^k), u^k - \bar{u}^k \rangle \geq \eta_k \left(F_k(x^k) - F_k(\bar{x}^k) + \frac{\gamma_k}{2} \|x^k - \bar{x}^k\|^2 \right)$$
$$+ \eta_k \left(E_k(x^k) - E_k(\bar{x}^k) - \frac{L_k}{2\kappa_k} \|\check{x}^k - x^k\|^2 \right)$$
$$+ \eta_k \left(G_k^*(y^k) - G_k^*(\bar{y}^k) + \frac{\rho_k}{2} \|y^k - \bar{y}^k\|^2 \right)$$
$$- \eta_k \langle K_k^* y^k, \bar{x}^k \rangle + \eta_k \langle K_k x^k, \bar{y}^k \rangle$$
$$(2.9) \qquad = \frac{1}{2} \llbracket u^k - \bar{u}^k \rrbracket_{\Gamma_k}^2 + \mathcal{G}_k^H(u^k, \bar{u}^k) - \frac{1}{2} \|u^k - \check{u}^k\|_{\eta_k \Omega_k}.$$

Following the testing methodology [10, 31], we apply the linear "testing operator" $\langle \cdot, u^k - \bar{u}^k \rangle_{\eta_k}$ to both sides of (2.4). This followed by (2.9) yields

$$0 \geq \langle u^k - \check{u}^k, u^k - \bar{u}^k \rangle_{\eta_k M_k} + \frac{1}{2} \llbracket u^k - \bar{u}^k \rrbracket_{\Gamma_k}^2 + \mathcal{G}_k^H(u^k, \bar{u}^k) - \frac{1}{2} \|u^k - \check{u}^k\|_{\eta_k \Omega_k} \quad (k = 1, \ldots, N).$$

Pythagoras' identity for the inner product and norm with respect to the operator $\eta_k M_k$ now yields

$$\frac{1}{2} \|\check{u}^k - \bar{u}^k\|_{\eta_k M_k}^2 \geq \frac{1}{2} \|u^k - \bar{u}^k\|_{\eta_k M_k + \Gamma_k}^2 + \mathcal{G}_k^H(u^k, \bar{u}^k) + \frac{1}{2} \|u^k - \check{u}^k\|_{\eta_k(M_k - \Omega_k)}^2 \quad (k = 1, \ldots, N).$$

Summing this over $k \in \{1, \ldots, N\}$, we obtain the claim.   □

We are now almost ready to state our main result regarding the dynamic regret of our algorithm. To proceed, we define the function $\mathring{G}_{1:N}$ by

$$\mathring{G}_{1:N}(z^{1:N}) := \sup_{\tilde{y}^{1:N} \in \mathcal{Y}_{1:N}} \left[\langle z^{1:N}, \tilde{y}^{1:N} \rangle - G_{1:N}^*(\tilde{y}^{1:N}) \right].$$

If the dual comparison sets $\mathcal{Y}_{1:N}$ were convex, then, recalling the formula $(f_1 + f_2)^* = f_1^* \square f_2^*$ for infimal convolutions (denoted $\square$) of convex functions f_1 and f_2, we would have $\mathring{G}_{1:N} = G_{1:N} \square \delta_{\mathcal{Y}_{1:N}}^*$. In general, $\mathring{G}_{1:N} \leq G_{1:N} \square \delta_{\mathcal{Y}_{1:N}}^*$.

It is worth noting that under suitable assumption on the comparison sequence, $\check{G} = \mathring{G}$; see [33, Example 3.4]. Moreover, if $\mathcal{Y}_{1:N} = \prod_{k=1}^N Y_k$, or even just $\mathcal{Y}_{1:N} \supset \operatorname{dom} G_{1:N}^*$, then it is also clear that $\mathring{G}_{1:N} = G_{1:N}$. In this case the next theorem provides a dynamic regret bound with respect to the original static objective $Q_{1:N} + G_{1:N} \circ K_{1:N}$ for the first N frames. Otherwise, providing bounds on $Q_{1:N} + \mathring{G}_{1:N} \circ K_{1:N}$, it modifies the objective by "sub-infimally" convolving $G_{1:N}$ with the temporal evolution constraints presented by the set of dual comparison sequences $\mathcal{Y}_{1:N}$. Typically $G_{1:N} \circ K_{1:N}$

would be a sum of independent static regulariser for the temporal frames 1 to N, for example, a sum of independent total variation terms for each frame. Then $\mathring{G}_{1:N}$ would be a temporally convolved total variation regulariser.

Compared to the "G-banana" $\check{G}_{1:N}$ of [33], the "G-doughnut" $\mathring{G}_{1:N}$ has a significantly simpler structure. Moreover, the following new result is symmetric, whereas the previous results of [33] were unsymmetric, employing $\check{G}_{1:N}$ for the iterates $x^{1:N}$, and the original $G_{1:N}$ for the comparison sequences $\bar{x}^{1:N}$. Generally, for the right hand side of the next main regret estimate to be meaningful, it is necessary for the set of comparison sequences to be bounded.

Theorem 2.6. *Let $N \geq 1$, and suppose Assumption 2.1 hold for $u^{1:N}$ generated by Algorithm 1 for an initial $u^0 \in X_0 \times Y_0$. Then*

$$(2.10) \quad [Q_{1:N}(x^{1:N}) + \mathring{G}_{1:N}(K_{1:N}x^{1:N})] - \sup_{\bar{x}^{1:N} \in \mathcal{X}_{1:N}} [Q_{1:N}(\bar{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N})]$$

$$\leq \sup_{\bar{u}^{0:N} \in \mathcal{U}_{0:N}} \left(\frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0 M_0 + \Gamma_0} + c_N(\bar{x}^{1:N}, y^{1:N}) + e_N(u^{0:N-1}, \bar{u}^{0:N}) \right),$$

where the prediction error $e_N(u^{0:N-1}, \bar{u}^{0:N})$ is given by (2.8), and the comparison set solution discrepancy

$$c_N(\bar{x}^{1:N}, y^{1:N}) := \inf_{\tilde{y}^{1:N} \in \mathcal{Y}_{1:N}} \langle K_{1:N}\bar{x}^{1:N}, y^{1:N} - \tilde{y}^{1:N} \rangle + G^*_{1:N}(\tilde{y}^{1:N}) - G^*_{1:N}(y^{1:N}).$$

Proof. Let $\bar{u}^{1:N} \in \mathcal{U}_{1:N}$. For any $\tilde{y}^{1:N} \in \mathcal{Y}_{1:N}$, by the definition of $\mathring{G}_{1:N}$ as a Fenchel conjugate of $G^*_{1:N} + \delta_{\mathcal{Y}_{1:N}}$, we have

$$(2.11) \quad \langle K_{1:N}\bar{x}^{1:N}, y^{1:N} \rangle - G^*_{1:N}(y^{1:N}) = \inf_{\tilde{y}^{1:N} \in \mathcal{Y}_{1:N}} \left(\langle K_{1:N}\bar{x}^{1:N}, y^{1:N} - \tilde{y}^{1:N} \rangle - G^*_{1:N}(y^{1:N}) + G^*_{1:N}(\tilde{y}^{1:N}) \right)$$

$$+ \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N})$$

$$= c_N(\bar{x}^{1:N}, y^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N}).$$

With this, we rearrange

$$\sum_{k=1}^{N} \mathcal{G}_k^H(u^k, \bar{u}^k) = \sum_{k=1}^{N} \eta_k \Big(\big[F_k(x^k) + E_k(x^k) + \langle K_k x^k, \bar{y}^k \rangle - G^*_k(\bar{y}^k) \big]$$

$$- \big[F_k(\bar{x}^k) + E_k(\bar{x}^k) + \langle K^*_k y^k, \bar{x}^k \rangle - G^*_k(y^k)) \big] \Big)$$

$$= \big[Q_{1:N}(x^{1:N}) + \langle K_{1:N}x^{1:N}, \bar{y}^{1:N} \rangle - G^*_{1:N}(\bar{y}^{1:N}) \big]$$

$$- \big[Q_{1:N}(\bar{x}^{1:N}) + \langle K_{1:N}\bar{x}^{1:N}, y^{1:N} \rangle - G^*_{1:N}(y^{1:N}) \big]$$

$$= \big[Q_{1:N}(x^{1:N}) + \langle K_{1:N}x^{1:N}, \bar{y}^{1:N} \rangle - G^*_{1:N}(\bar{y}^{1:N}) \big]$$

$$- \big[Q_{1:N}(\bar{x}^{1:N}) + c_N(\bar{x}^{1:N}, y^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N}) \big].$$

Using this equation in the claim of Lemma 2.5 readily establishes that

$$(2.12) \quad D_N(\bar{u}^{1:N}) \leq \frac{1}{2}\|u^0 - \bar{u}^0\|^2_{\eta_0 M_0 + \Gamma_0} + c_N(\bar{x}^{1:N}, y^{1:N}) + e_N(u^{0:N-1}, \bar{u}^{0:N}).$$

for

$$D_N(\bar{u}^{1:N}) := Q_{1:N}(x^{1:N}) + \langle K_{1:N}x^{1:N}, \bar{y}^{1:N} \rangle - G^*_{1:N}(\bar{y}^{1:N}) - Q_{1:N}(\bar{x}^{1:N}) - \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N}).$$

We have

$$
\begin{aligned}
\sup_{\bar{u}^{0:N} \in \mathcal{U}_{0:N}} D_N(\bar{u}^{1:N}) &= \sup_{\bar{u}^{1:N} \in \mathcal{U}_{1:N}} D_N(\bar{u}^{1:N}) \\
&\geq \sup_{\bar{u}^{1:N} \in \mathcal{U}_{1:N}} \left(Q_{1:N}(x^{1:N}) + \langle K_{1:N} x^{1:N}, \bar{y}^{1:N} \rangle - G_{1:N}^*(\bar{y}^{1:N}) \right) \\
&\quad - \sup_{\bar{u}^{1:N} \in \mathcal{U}_{1:N}} \left(Q_{1:N}(\bar{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N} \bar{x}^{1:N}) \right) \\
&= Q_{1:N}(x^{1:N}) + \mathring{G}_{1:N}(K_{1:N} x^{1:N}) - \sup_{\bar{x}^{1:N} \in X_{1:N}} \left(Q_{1:N}(\bar{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N} \bar{x}^{1:N}) \right).
\end{aligned}
$$

Therefore, the claim follows by taking the supremum over $\bar{u}^{0:N} \in \mathcal{U}_{0:N}$ in (2.12). □

Remark 2.7 (Comparison set solution discrepancy). The comparison set solution discrepancy c_N is the price for symmetricity in Theorem 2.6, as compared to [33]. It is non-positive and therefore can be made to disappear from (2.10) if

$$(2.13) \qquad G_{1:N}(K_{1:N} \bar{x}^{1:N}) \leq \langle K_{1:N} \bar{x}^{1:N}, \tilde{y}^{1:N} \rangle - G_{1:N}^*(\tilde{y}^{1:N}) \quad \text{for some} \quad \tilde{y}^{1:N} \in \mathcal{Y}_{1:N},$$

since the definition of the Fenchel biconjugate establishes $\langle K_{1:N} \bar{x}^{1:N}, y^{1:N} \rangle - G_{1:N}^*(y^{1:N}) \leq G_{1:N}(K_{1:N} \bar{x}^{1:N})$. The condition (2.13) also establishes

$$\mathring{G}_{1:N}(K_{1:N} \bar{x}^{1:N}) = G_{1:N}(K_{1:N} \bar{x}^{1:N}).$$

Alternatively, if $\mathcal{Y}_{1:N}$ is large enough, it may be possible to make c_N small. In particular, if the algorithm-generated iterates $y^{1:N}$ are in $\mathcal{Y}_{1:N}$, we have $c_N(\bar{x}^{1:N}, y^{1:N}) = 0$. In our applications of interest, this seems, however, unlikely. Instead, in the following Section 3, we will relate (2.13) to total variation preserving predictors and true temporal couplings, taking $\tilde{y}^{1:N}$ such that $(\bar{x}^{1:N}, \tilde{y}^{1:N}) \in \mathcal{U}_{1:N}$.

3 PSEUDO–AFFINE PREDICTORS

Our first purpose in this section is to estimate in Section 3.1 the prediction errors e_N, defined in (2.8), for a class of what we call *pseudo-affine predictors*. We then provide examples of such predictors in Sections 3.2 to 3.4, based on enforcing the preservation of salient relationships between the primal and dual variables. Roughly speaking, these pointwise total variation and angle-preserving predictors model the fact that new shapes, hence significant changes in local total variation or angles, can only emerge near the image boundaries, i.e., in a set that has a small measure and hence a small effect on the prediction error.

In Algorithm 1, the time evolution of primal and dual variables is described by the predictors $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$ where we wrote $(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k)$ for all $k \in \mathbb{N}$. Throughout this section, we assume that

$$(3.1) \qquad P_k(x^k, y^k) = (W_k x^k + a_{k+1}, T_k y^k + b_{k+1})$$

for some $W_k \in \mathbb{L}(X_k; X_{k+1})$, $T_k \in \mathbb{L}(Y_k; Y_{k+1})$, $a_k \in X_{k+1}$, $b_k \in Y_{k+1}$. Note that although P_k is assumed affine, it may hide nonlinear dependencies through the dependence on the iteration k. This generality allows W_k or T_k to have nonlinear factors that also depend on the primal variable (as elucidated in the examples we consider for our purpose). We take the same form of primal-dual temporal coupling operators $\bar{P}_k$ for some $\bar{W}_k \in \mathbb{L}(X_k; X_{k+1})$, $\bar{T}_k \in \mathbb{L}(Y_k; Y_{k+1})$, $\bar{a}_{k+1} \in X_{k+1}$ and $\bar{b}_{k+1} \in Y_{k+1}$ to describe the evolution of variables in the comparison sequence. More succinctly, for some family $\mathcal{P}$ of sequences of primal-dual temporal coupling operators $\{\tilde{P}_k\}_{k=0}^{\infty}$, we define the set of comparison sequences

$$\mathcal{U} := \left\{ \bar{u}^{0:\infty} \in \prod_{k=0}^{\infty} X_k \times Y_k \mid \bar{u}^0 = (\bar{x}^0, \bar{y}^0) \in \mathcal{U}_0, \bar{u}^{k+1} = \tilde{P}_k(\bar{x}^k, \bar{y}^k), (\tilde{P}_k)_{k=0}^{\infty} \in \mathcal{P} \right\}.$$

3.1 PREDICTION BOUNDS AND PENALTIES

Given the general form (3.1) of primal-dual predictors and temporal coupling operators, we first derive respective bounds for primal and dual predictions. This entails computing (Lipschitz-like) factors $\Lambda_k, \Theta_k > 0$ and prediction penalties $\varepsilon_{k+1}, \tilde{\varepsilon}_{k+1} \geq 0$ as in the next lemma.

Lemma 3.1. *For all $k \in \mathbb{N}$ and for some $M_{\bar{x}}, M_{\bar{y}} > 0$, let $\bar{x}^k \in X_k$ with $\|\bar{x}^k\|_{X_k}^2 \leq M_{\bar{x}}$, $\bar{y}^k \in Y_k$ with $\|\bar{y}^k\|_{Y_k}^2 \leq M_{\bar{y}}$,*

$$(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k) = \left(W_k x^k + a_{k+1}, T_k y^k + b_{k+1}\right), \quad \text{and}$$
$$(\bar{x}^{k+1}, \bar{y}^{k+1}) = \bar{P}_k(\bar{x}^k, \bar{y}^k) = \left(\bar{W}_k \bar{x}^k + \bar{a}_{k+1}, \bar{T}_k \bar{y}^k + \bar{b}_{k+1}\right)$$

for a fixed $\{\bar{P}_k\}_{k=0}^{\infty} \in \mathcal{P}$, $W_k, \bar{W}_k \in \mathbb{L}(X_k; X_{k+1})$, $T_k, \bar{T}_k \in \mathbb{L}(Y_k; Y_{k+1})$, $a_{k+1}, \bar{a}_{k+1} \in X_{k+1}$, and $b_{k+1}, \bar{b}_{k+1} \in Y_{k+1}$. Then for any $\pi_k > 0$ and $\Lambda_k > \|W_k\|^2$,

$$\frac{1}{2}\|\check{x}^{k+1} - \bar{x}^{k+1}\|_{X_{k+1}}^2 \leq \frac{\Lambda_k}{2}\|x^k - \bar{x}^k\|_{X_k}^2 + \varepsilon_{k+1}$$

where

$$\varepsilon_{k+1} := \frac{\Lambda_k M_{\bar{x}}(1 + \pi_k)}{\Lambda_k - \|W_k\|^2}\|W_k - \bar{W}_k\|^2 + \frac{\Lambda_k(1 + \pi_k^{-1})}{\Lambda_k - \|W_k\|^2}\|a_{k+1} - \bar{a}_{k+1}\|^2.$$

Similarly, for any $\tilde{\pi}_k > 0$ and $\Theta_k > \|T_k\|^2$,

$$\frac{1}{2}\|\check{y}^{k+1} - \bar{y}^{k+1}\|_{Y_{k+1}}^2 \leq \frac{\Theta_k}{2}\|y^k - \bar{y}^k\|_{Y_k}^2 + \tilde{\varepsilon}_{k+1}$$

where

$$\tilde{\varepsilon}_{k+1} := \frac{\Theta_k M_{\bar{y}}(1 + \tilde{\pi}_k)}{\Theta_k - \|T_k\|^2}\|T_k - \bar{T}_k\|^2 + \frac{\Theta_k(1 + \tilde{\pi}_k^{-1})}{\Theta_k - \|T_k\|^2}\|b_{k+1} - \bar{b}_{k+1}\|^2.$$

Proof. For any $t > 0$, we apply Young's inequality to obtain

$$\begin{aligned}
\|\check{x}^{k+1} - \bar{x}^{k+1}\|^2 &= \|W_k x^k + a_{k+1} - \bar{W}_k \bar{x}^k - \bar{a}_{k+1}\|^2 \\
&\leq (1+t)\|W_k x^k - W_k \bar{x}^k\|^2 + (1+t^{-1})\|W_k \bar{x}^k - \bar{W}_k \bar{x}^k + a_{k+1} - \bar{a}_{k+1}\|^2 \\
&\leq (1+t)\|W_k\|^2\|x^k - \bar{x}^k\|^2 \\
&\quad + (1+t^{-1})\left[(1+\pi_k)\|(W_k - \bar{W}_k)(\bar{x}^k)\|^2 + (1+\pi_k^{-1})\|a_{k+1} - \bar{a}_{k+1}\|^2\right] \\
&\leq (1+t)\Lambda_k\|x^k - \bar{x}^k\|^2 \\
&\quad + (1+t^{-1})\left[(1+\pi_k)\|W_k - \bar{W}_k\|^2 M_{\bar{x}} + (1+\pi_k^{-1})\|a_{k+1} - \bar{a}_{k+1}\|^2\right].
\end{aligned}$$

Choosing $t = 1 - \|W_k\|^2/\Lambda_k$ yields the desired Lipschitz-like constants and penalties for the primal prediction. The dual prediction bounds and penalties are obtained by a similar computation. $\square$

We can now state our main result regarding the prediction errors e_N. The following theorem also imposes additional step length condition so that the prediction errors remain bounded.

Theorem 3.2. *Let Assumption 2.1 together with the assumptions of Lemma 3.1 hold. Suppose further that $\|\frac{\eta_k}{\eta_{k+1}} K_k - T_k^* K_{k+1} W_k\|^2 \leq C_k$ for some $C_k \geq 0$. If the testing and step length parameters satisfy*

$$(3.2) \qquad\qquad \varphi_k(1 + \gamma_k \tau_k) > \varphi_{k+1}\Lambda_k,$$

then

$$e_N(u^{0:N-1}, \bar{u}^{0:N}) \leq \sum_{k=0}^{N-1} \varepsilon_{k+1}^{\dagger}(\bar{u}^k),$$

where, for any $\pi_k, \tilde{\pi}_k, \beta > 0$ and $\kappa \in (0, 1)$,

(3.3)
$$
\begin{aligned}
\varepsilon_{k+1}^{\dagger}(\bar{u}^k) = {}& \left(\frac{\psi_{k+1}\Theta_k - \kappa\psi_k(1+\rho_k\sigma_k)}{2} + \frac{\eta_{k+1}^2(C_k\beta + \|K_{k+1}\|^2\|W_k\|^2)}{2\beta(\varphi_k(1+\gamma_k\tau_k) - \varphi_{k+1}\Lambda_k)} \right) \|y^k - \bar{y}^k\|^2 \\
& + \left(\frac{\eta_{k+1}^2\|K_{k+1}\|^2\|T_k\|^2}{2(1-\kappa)\psi_k(1+\rho_k\sigma_k)} + \frac{\eta_{k+1}\|K_{k+1}\|^2}{2} + \frac{\varphi_{k+1}\Lambda_k}{\Lambda_k - \|W_k\|^2} \right) \\
& \qquad \cdot \left[M_{\bar{x}}(1+\pi_k)\|\bar{W}_k - W_k\|^2 + (1+\pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2 \right] \\
& + \left(\frac{\eta_{k+1}^2(C_k\beta + \|K_{k+1}\|^2\|W_k\|^2)}{2(\varphi_k(1+\gamma_k\tau_k) - \varphi_{k+1}\Lambda_k)} + \frac{\eta_{k+1}}{2} + \frac{\psi_{k+1}\Theta_k}{\Theta_k - \|T_k\|^2} \right) \\
& \qquad \cdot \left[M_{\bar{y}}(1+\tilde{\pi}_k)\|\bar{T}_k - T_k\|^2 + (1+\tilde{\pi}_k^{-1})\|\bar{b}_{k+1} - b_{k+1}\|^2 \right].
\end{aligned}
$$

Remark 3.3. The first line of (3.3) depends on the boundedness of the dual iterates; the remaining lines depend on the difference between the predictions and true temporal couplings, which disappear whenever the predictions are as good as the true temporal couplings. The factors C_k measure the compatibility of the predictions with the operators K_k. In view of Theorem 3.2, the regret estimate of Theorem 2.6 now shows that

$$
[Q_{1:N}(x^{1:N}) + \mathring{G}_{1:N}(K_{1:N}x^{1:N})] - \sup_{\bar{x}^{1:N} \in \mathcal{X}_{1:N}} [Q_{1:N}(\bar{x}^{1:N}) + \mathring{G}_{1:N}(K_{1:N}\bar{x}^{1:N})]
$$

$$
\leq \sup_{\bar{u}^{0:N} \in \mathcal{U}_{0:N}} \left(\frac{1}{2}\|u^0 - \bar{u}^0\|_{\eta_0 M_0 + \Gamma_0}^2 + c_N(\bar{x}^{1:N}, y^{1:N}) + \sum_{k=0}^{N-1} \varepsilon_{k+1}^{\dagger}(\bar{u}^k) \right).
$$

Minding the interpretation of $c_N(\bar{x}^{1:N}, y^{1:N})$ in Remark 2.7, the regret therefore depends on how well (a) the comparison set solves the static problems, and (b) the predictors track the true temporal couplings.

In the unaccelerated setting of Example 2.2, with linear temporal evolution and predictors ($a_{k+1} = \bar{a}_{k+1} = 0$, and $b_{k+1} = \bar{b}_{k+1} = 0$), we can simplify

$$
\begin{aligned}
\varepsilon_{k+1}^{\dagger}(\bar{u}^k) = {}& \left(\frac{\tau}{\sigma}\frac{\Theta_k - \kappa(1+\rho_k\sigma)}{2} + \frac{\tau^2(C_k\beta + \|K_{k+1}\|^2\|W_k\|^2)}{2\beta(1+\gamma_k\tau - \Lambda_k)} \right) \|y^k - \bar{y}^k\|^2 \\
& + \left(\frac{\tau\sigma\|K_{k+1}\|^2\|T_k\|^2}{2(1-\kappa)(1+\rho_k\sigma)} + \frac{\tau\|K_{k+1}\|^2}{2} + \frac{\Lambda_k}{\Lambda_k - \|W_k\|^2} \right) \cdot M_{\bar{x}}\|\bar{W}_k - W_k\|^2 \\
& + \left(\frac{\tau^2(C_k\beta + \|K_{k+1}\|^2\|W_k\|^2)}{2(1+\gamma_k\tau - \Lambda_k)} + \frac{\tau}{2} + \frac{\tau}{\sigma}\frac{\Theta_k}{\Theta_k - \|T_k\|^2} \right) \cdot M_{\bar{y}}\|\bar{T}_k - T_k\|^2.
\end{aligned}
$$

Moreover, if we have sufficient strong convexity compared to the bounds Θ_k and Λ_k, i.e., such that $1 + \rho_k\sigma \geq \kappa^{-1}\Theta_k$ and $1 + \gamma_k\tau > \Lambda_k$, we estimate

$$
\begin{aligned}
\varepsilon_{k+1}^{\dagger}(\bar{u}^k) \leq {}& \left(\frac{\tau\sigma\|K_{k+1}\|^2\|T_k\|^2}{2(1-\kappa)(1+\rho_k\sigma)} + \frac{\tau\|K_{k+1}\|^2}{2} + \frac{\Lambda_k}{\Lambda_k - \|W_k\|^2} \right) \cdot M_{\bar{x}}\|\bar{W}_k - W_k\|^2 \\
& + \left(\frac{\tau}{2} + \frac{\tau}{\sigma}\frac{\Theta_k}{\Theta_k - \|T_k\|^2} \right) \cdot M_{\bar{y}}\|\bar{T}_k - T_k\|^2.
\end{aligned}
$$

Additionally, using $\tau\sigma\|K_{k+1}\|^2 \leq 1$ from (2.2b) in Assumption 2.1, we further estimate

$$
\begin{aligned}
\varepsilon_{k+1}^{\dagger}(\bar{u}^k) \leq {}& \left(\frac{\|T_k\|^2}{2(1-\kappa)\Theta_k} + \frac{1}{2\sigma} + \frac{\Lambda_k}{\Lambda_k - \|W_k\|^2} \right) \cdot M_{\bar{x}}\|\bar{W}_k - W_k\|^2 \\
& + \frac{\tau}{\sigma} \left(\frac{1}{2\sigma} + \frac{\Theta_k}{\Theta_k - \|T_k\|^2} \right) \cdot M_{\bar{y}}\|\bar{T}_k - T_k\|^2.
\end{aligned}
$$

Thus, in that case, the prediction penalty mainly depends on

(a) the bounds $M_{\bar{x}} \geq \|\bar{x}^k\|^2$ and $M_{\bar{y}} \geq \|\bar{y}^k\|^2$ on the comparison sequences;

(b) the closeness of the prediction operators T_k and W_k to the true temporal couplings $\bar{T}_k$ and $\bar{W}_k$; and

(c) the looseness of the upper bounds Θ_k and Λ_k on $\|T_k\|$ and $\|W_k\|$.

Proof of Theorem 3.2. By the definition of the primal and dual predictors, we have

$$
(3.4) \quad \|\check{u}^{k+1} - \bar{u}^{k+1}\|^2_{\eta_{k+1} M_{k+1}} = \varphi_{k+1} \|x^{k+1} - \check{x}^{k+1}\|^2 + \psi_{k+1} \|y^{k+1} - \check{y}^{k+1}\|^2
$$
$$
- 2\eta_{k+1} \langle K_{k+1}(W_k x^k - \bar{W}_k \bar{x}^k + a_{k+1} - \bar{a}_{k+1}), T_k y^k - \bar{T}_k \bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle
$$

and

$$
\|u^k - \bar{u}^k\|^2_{\eta_k M_k + \Gamma_k} = \varphi_k(1 + \gamma_k \tau_k)\|x^k - \bar{x}^k\|^2 + \psi_k(1 + \rho_k \sigma_k)\|y^k - \bar{y}^k\|^2
$$
$$
(3.5) \quad\quad\quad - 2\eta_k \langle K_k(x^k - \bar{x}^k), y^k - \bar{y}^k \rangle.
$$

Expanding the last term at the right-hand side of (3.4) (without the scalar factor η_{k+1}) yields

$$
-2 \langle K_{k+1}(W_k x^k - \bar{W}_k \bar{x}^k + a_{k+1} - \bar{a}_{k+1}), T_k y^k - \bar{T}_k \bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle
$$
$$
= -2 \langle K_{k+1} W_k(x^k - \bar{x}^k), T_k(y^k - \bar{y}^k) \rangle
$$
$$
+ 2 \langle K_{k+1} W_k(\bar{x}^k - x^k), (T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle
$$
$$
+ 2 \langle K_{k+1}(\bar{W}_k - W_k)\bar{x}^k + K_{k+1}(\bar{a}_{k+1} - a_{k+1}), T_k(y^k - \bar{y}^k) \rangle
$$
$$
(3.6) \quad\quad + 2 \langle K_{k+1}(\bar{W}_k - W_k)\bar{x}^k + K_{k+1}(\bar{a}_{k+1} - a_{k+1}), (T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle.
$$

Using Young's inequality for some $s > 0$,

$$
2 \langle K_{k+1} W_k(\bar{x}^k - x^k), (T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle
$$
$$
\leq s\|K_{k+1}\|^2 \|W_k\|^2 \|\bar{x}^k - x^k\|^2 + s^{-1}\|(T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1}\|^2
$$
$$
\leq s\|K_{k+1}\|^2 \|W_k\|^2 \|\bar{x}^k - x^k\|^2 + s^{-1}(1 + \tilde{\pi}_k)M_{\bar{y}}\|T_k - \bar{T}_k\|^2
$$
$$
(3.7) \quad\quad + s^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2.
$$

Similarly, by Young's inequality (for some $r, \zeta > 0$) we obtain

$$
2 \langle K_{k+1}(\bar{W}_k - W_k)\bar{x}^k + K_{k+1}(\bar{a}_{k+1} - a_{k+1}), T_k(y^k - \bar{y}^k) \rangle
$$
$$
= 2 \langle (\bar{W}_k - W_k)\bar{x}^k + \bar{a}_{k+1} - a_{k+1}, K_{k+1}^* T_k(y^k - \bar{y}^k) \rangle
$$
$$
\leq r\|(\bar{W}_k - W_k)\bar{x}^k + \bar{a}_{k+1} - a_{k+1}\|^2 + r^{-1}\|K_{k+1}\|^2 \|T_k\|^2 \|y^k - \bar{y}^k\|^2
$$
$$
\leq M_{\bar{x}} r(1 + \pi_k)\|\bar{W}_k - W_k\|^2 + r(1 + \pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2
$$
$$
(3.8) \quad\quad + r^{-1}\|K_{k+1}\|^2 \|T_k\|^2 \|y^k - \bar{y}^k\|^2
$$

and

$$
(3.9) \quad 2 \langle K_{k+1}(\bar{W}_k - W_k)\bar{x}^k + K_{k+1}(\bar{a}_{k+1} - a_{k+1}), (T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1} \rangle
$$
$$
\leq \zeta\|K_{k+1}(\bar{W}_k - W_k)\bar{x}^k + K_{k+1}(\bar{a}_{k+1} - a_{k+1})\|^2
$$
$$
+ \zeta^{-1}\|(T_k - \bar{T}_k)\bar{y}^k + b_{k+1} - \bar{b}_{k+1}\|^2
$$
$$
\leq \zeta M_{\bar{x}}(1 + \pi_k)\|K_{k+1}\|^2 \|\bar{W}_k - W_k\|^2
$$
$$
+ \zeta(1 + \pi_k^{-1})\|K_{k+1}\|^2 \|\bar{a}_{k+1} - a_{k+1}\|^2
$$
$$
+ \zeta^{-1} M_{\bar{y}}(1 + \tilde{\pi}_k)\|T_k - \bar{T}_k\|^2 + \zeta^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2.
$$

Substituting (3.7), (3.8) and (3.9) into (3.6) yields

$$-2\langle K_{k+1}(W_k x^k - \bar{W}_k \bar{x}^k), T_k y^k - \bar{T}_k \bar{y}^k \rangle$$
$$\leq -2\langle K_{k+1} W_k(x^k - \bar{x}^k), T_k(y^k - \bar{y}^k)\rangle$$
$$+ s\|K_{k+1}\|^2\|W_k\|^2\|\bar{x}^k - x^k\|^2 + s^{-1}(1 + \tilde{\pi}_k)M_{\bar{y}}\|T_k - \bar{T}_k\|^2$$
$$+ s^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2 + M_{\bar{x}}r(1 + \pi_k)\|\bar{W}_k - W_k\|^2$$
$$+ r(1 + \pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2 + r^{-1}\|K_{k+1}\|^2\|T_k\|^2\|y^k - \bar{y}^k\|^2$$
$$+ \zeta M_{\bar{x}}(1 + \pi_k)\|K_{k+1}\|^2\|\bar{W}_k - W_k\|^2 + \zeta(1 + \pi_k^{-1})\|K_{k+1}\|^2\|\bar{a}_{k+1} - a_{k+1}\|^2$$
$$+ \zeta^{-1}M_{\bar{y}}(1 + \tilde{\pi}_k)\|T_k - \bar{T}_k\|^2 + \zeta^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2.$$

Combining this inequality with (3.4), then invoking the prediction bounds and penalties from Lemma 3.1, we get

$$\|\check{u}^{k+1} - \bar{u}^{k+1}\|_{\eta_{k+1}M_{k+1}}^2 \leq \varphi_{k+1}\Lambda_k\|x^k - \bar{x}^k\|^2 + \psi_{k+1}\Theta_k\|y^k - \bar{y}^k\|^2 + \eta_{k+1}h_{k+1}$$
(3.10)
$$+ 2\varphi_{k+1}\varepsilon_{k+1} + 2\psi_{k+1}\tilde{\varepsilon}_{k+1}$$

for

$$h_{k+1} := -2\langle K_{k+1}W_k(x^k - \bar{x}^k), T_k(y^k - \bar{y}^k)\rangle + s\|K_{k+1}\|^2\|W_k\|^2\|\bar{x}^k - x^k\|^2$$
$$+ s^{-1}(1 + \tilde{\pi}_k)M_{\bar{y}}\|T_k - \bar{T}_k\|^2 + s^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2$$
$$+ M_{\bar{x}}r(1 + \pi_k)\|\bar{W}_k - W_k\|^2 + r(1 + \pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2$$
$$+ r^{-1}\|K_{k+1}\|^2\|T_k\|^2\|y^k - \bar{y}^k\|^2 + \zeta\|K_{k+1}\|^2 M_{\bar{x}}(1 + \pi_k)\|\bar{W}_k - W_k\|^2$$
$$+ \zeta\|K_{k+1}\|^2(1 + \pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2 + \zeta^{-1}M_{\bar{y}}(1 + \tilde{\pi}_k)\|T_k - \bar{T}_k\|^2$$
$$+ \zeta^{-1}(1 + \tilde{\pi}_k^{-1})\|b_{k+1} - \bar{b}_{k+1}\|^2.$$

Furthermore, we note that by another application of Young's inequality (for some $t > 0$) and by using the assumption that $\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\|^2 \leq C_k$, it follows that

(3.11)
$$2\eta_k\langle K_k(x^k - \bar{x}^k), y^k - \bar{y}^k\rangle - 2\eta_{k+1}\langle K_{k+1}W_k(x^k - \bar{x}^k), T_k(y^k - \bar{y}^k)\rangle$$
$$= 2\eta_{k+1}\left\langle \left(\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right)(x^k - \bar{x}^k), y^k - \bar{y}^k\right\rangle$$
$$\leq \eta_{k+1}C_k t\|x^k - \bar{x}^k\|^2 + t^{-1}\eta_{k+1}\|y^k - \bar{y}^k\|^2$$

Subtracting (3.5) from (3.10), and using (3.11) yields

(3.12)
$$\|\check{u}^{k+1} - \bar{u}^{k+1}\|_{\eta_{k+1}M_{k+1}}^2 - \|u^k - \bar{u}^k\|_{\eta_k M_k + \Gamma_k}^2$$
$$\leq \left(\varphi_{k+1}\Lambda_k + \eta_{k+1}C_k t + \eta_{k+1}\|K_{k+1}\|^2\|W_k\|^2 s - \varphi_k(1 + \gamma_k\tau_k)\right)\|\bar{x}^k - x^k\|^2$$
$$+ \left(\psi_{k+1}\Theta_k + \eta_{k+1}t^{-1} + \eta_{k+1}\|K_{k+1}\|^2\|T_k\|^2 r^{-1} - \psi_k(1 + \rho_k\sigma_k)\right)\|y^k - \bar{y}^k\|^2$$
$$+ \eta_{k+1}\left(r + \|K_{k+1}\|^2\zeta\right)\left[M_{\bar{x}}(1 + \pi_k)\|\bar{W}_k - W_k\|^2 + (1 + \pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2\right]$$
$$+ \eta_{k+1}\left(s^{-1} + \zeta^{-1}\right)\left[M_{\bar{y}}(1 + \tilde{\pi}_k)\|\bar{T}_k - T_k\|^2 + (1 + \tilde{\pi}_k^{-1})\|\bar{b}_{k+1} - b_{k+1}\|^2\right]$$
$$+ 2\varphi_{k+1}\varepsilon_{k+1} + 2\psi_{k+1}\tilde{\varepsilon}_{k+1}.$$

Since $\varphi_k(1 + \gamma_k\tau_k) - \varphi_{k+1}\Lambda_k > 0$ by assumption, setting $t = \beta s$ for some $\beta > 0$ and choosing

$$s = \frac{\varphi_k(1 + \gamma_k\tau_k) - \varphi_{k+1}\Lambda_k}{\eta_{k+1}(C_k\beta + \|K_{k+1}\|^2\|W_k\|^2)} > 0$$

gives us

$$\varphi_{k+1}\Lambda_k + \eta_{k+1}C_k t + \eta_{k+1}\|K_{k+1}\|^2\|W_k\|^2 s - \varphi_k(1 + \gamma_k\tau_k) = 0.$$

By using the definition of ε_{k+1} and $\tilde{\varepsilon}_{k+1}$ from Lemma 3.1, and by setting

$$r = \frac{\eta_{k+1}\|K_{k+1}\|^2\|T_k\|^2}{\psi_k(1-\kappa)(1+\rho_k\sigma_k)} \quad \text{and} \quad \zeta = 1,$$

the remaining terms on the right-hand side of (3.12) give the prediction penalties. Thus, (3.12) turns into the desired overall prediction bound. $\qquad \square$

Remark 3.4. Alternatively, for any $\beta, \mu, \omega \in (0,1)$, we can set

$$t = \frac{\eta_{k+1}}{(1-\omega)(1+\rho_k\sigma_k)\psi_k}, \qquad\qquad s = \frac{(1-\mu)\varphi_k(1+\gamma_k\tau_k)}{\eta_{k+1}\|K_{k+1}\|^2\|W_k\|^2},$$

$$r = \frac{\eta_{k+1}\|K_k\|^2\|T_k\|^2}{(1-\beta)\omega\psi_k(1+\rho_k\sigma_k)}, \quad \text{and} \qquad\qquad \zeta = 1$$

in (3.12) while imposing the dual predictor restriction and the primal metric update bounds

$$\beta\kappa\psi_k(1+\rho_k\sigma_k) > \psi_{k+1}\Theta_k \quad \text{and}$$

$$\mu\varphi_k(1+\gamma_k\tau_k) > \varphi_{k+1}\Lambda_k + \frac{\eta_{k+1}^2 C_k}{\psi_k(1-\kappa)(1+\sigma_k\rho_k)}.$$

Then $e_N(u^{0:N-1}, \bar{u}^{0:N}) \leq \sum_{k=0}^{N-1} \varepsilon_{k+1}^{\dagger}(\bar{u}^k)$ with

$$
\begin{aligned}
\varepsilon_{k+1}^{\dagger} = &\left(\frac{\eta_{k+1}^2\|K_{k+1}\|^2\|T_k\|^2}{2(1-\beta)\kappa\omega\psi_k(1+\rho_k\sigma_k)} + \frac{\eta_{k+1}\|K_{k+1}\|^2}{2} + \frac{\varphi_{k+1}\Lambda_k}{\Lambda_k - \|W_k\|^2} \right) \\
&\cdot \left[M_{\bar{x}}(1+\pi_k)\|\bar{W}_k - W_k\|^2 + (1+\pi_k^{-1})\|\bar{a}_{k+1} - a_{k+1}\|^2 \right] \\
+ &\left(\frac{\eta_{k+1}^2\|K_{k+1}\|^2\|W_k\|^2}{2(1-\mu)\varphi_k(1+\gamma_k\tau_k)} + \frac{\eta_{k+1}}{2} + \frac{\psi_{k+1}\Theta_k}{\Theta_k - \|T_k\|^2} \right) \\
&\cdot \left[M_{\bar{y}}(1+\tilde{\pi}_k)\|\bar{T}_k - T_k\|^2 + (1+\tilde{\pi}_k^{-1})\|\bar{b}_{k+1} - b_{k+1}\|^2 \right].
\end{aligned}
$$

Observe that, in contrast to Theorem 3.2, this version of $\varepsilon_{k+1}^{\dagger}$ does not depend on $\|y^k - \bar{y}^k\|^2$.

3.2 TOTAL VARIATION PRESERVING PREDICTORS

We now consider examples of pseudo-linear predictors within the framework of optical flow. In what follows, we let

$$G_k = \alpha\|\cdot\|_{2,1} \text{ for some } \alpha > 0, \text{ and } K_k = D$$

some $D \in \mathbb{L}(X_k; Y_k)$, typically, but not necessarily a differential operator, with $Y_k \subset L^2(\Omega; \mathbb{R}^m)$ for some $m \geq 1$ and a domain Ω such that the "global 1-norm, pointwise 2-norm" is defined. We will make more specific assumptions on Ω and X_k for individual predictors. Recall our notation for the primal-dual predictor $P_k : X_k \times Y_k \to X_{k+1} \times Y_{k+1}$ where we wrote $(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k)$ for all $k \in \mathbb{N}$. We say that a predictor P_k is (strictly) *total variation preserving*[2] if the primal and dual predictions satisfy

(3.13) $\alpha\|Dx^k\|_{2,1} = \langle Dx^k, y^k \rangle$ and $\|y^k\|_{2,\infty} \leq \alpha$

$$\implies \alpha\|D\check{x}^{k+1}\|_{2,1} = \langle D\check{x}^{k+1}, \check{y}^{k+1} \rangle \quad \text{(and, for strict preservation, } \|\check{y}^{k+1}\|_{2,\infty} \leq \alpha\text{)}.$$

---

[2] The nomenclature changes if K_k is chosen differently from a differential operator.

We give several examples of this kind of predictor in the following discussion. In each of these examples, we establish conditions guaranteeing the assumptions of Theorem 3.2 on the predictors, that is, bounds $C_k > \|\frac{\eta_k}{\eta_{k+1}} K_k - T_k^* K_{k+1} W_k\|^2$, $\Theta_k > \|T_k\|^2$, and $\Lambda_k > \|W_k\|^2$, for the linear parts T_k and W_k of the primal and dual predictors, where the overall prediction $P_k(x^k, y^k) = (W_k x^k + a_k, T_k y^k + b_k)$.

Observe that (2.13), which guarantees that the comparison set solution discrepancy $c_N(\bar{x}^{1:N}, y^{1:N}) \leq 0$, holds if $\alpha\|D\bar{x}^0\|_{2,1} = \langle D\bar{x}^0, \bar{y}^0 \rangle$ and $\|\bar{y}^0\|_{2,\infty} \leq \alpha$, and the true temporal coupling is strictly total variation preserving:

$$(3.14) \quad \alpha\|D\bar{x}^k\|_{2,1} = \langle D\bar{x}^k, \bar{y}^k \rangle \text{ and } \|\bar{y}^k\|_{2,\infty} \leq \alpha$$
$$\implies \alpha\|D\bar{x}^{k+1}\|_{2,1} = \langle D\bar{x}^{k+1}, \bar{y}^{k+1} \rangle \text{ and } \|\bar{y}^{k+1}\|_{2,\infty} \leq \alpha.$$

POINTWISE PRESERVATION IN L^2

For a domain $\Omega \subset \mathbb{R}^n$, we work with the subspaces $X_k \subset L^2(\Omega)$, $Y_k \subset L^2(\Omega; \mathbb{R}^m)$, and a subset $\mathcal{V}_k \subset H^1(\Omega; \Omega)$ of bijective displacement fields, from which we take a measured v^k and a true $\bar{v}^k$ that give the primal prediction as $\check{x}^{k+1} = x^k \circ v^k$, and the true primal temporal evolution as $\bar{x}^{k+1} = \bar{x}^k \circ \bar{v}^k$.

Lemma 3.5. *Let $\mathcal{V}_k \subset H^1(\Omega; \Omega)$ be a set of bijective displacement fields satisfying $(v^k)^{-1} \in H^1(\Omega; \Omega)$ for all $v^k \in \mathcal{V}_k$, and*

$$\Lambda_{\mathcal{V}_k} := \sup_{v^k \in \mathcal{V}_k, \xi \in \Omega} |\det \nabla(v^k)^{-1}(\xi)| < \infty.$$

Let $x^k \in X_k$, $y^k \in Y_k$, $v^k \in \mathcal{V}_k$, and

$$(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k) = (W_k x^k, T_k y^k),$$

where the primal $W_k : X_k \to X_{k+1}$ and dual predictors $T_k : Y_k \to Y_{k+1}$ are defined pointwise by

$$(3.15a) \qquad (W_k x^k)(\xi) := (x^k \circ v^k)(\xi) \quad and \quad (T_k y^k)(\xi) := t^k(\xi) y^k(v^k(\xi))$$

with $t^k(\xi)$ given by

$$(3.15b) \qquad t^k(\xi) := \begin{cases} \frac{|\nabla v^k(\xi)^* D x^k(v^k(\xi))|}{|D x^k(v^k(\xi))|}(\nabla v^k(\xi))^{-1}, & |D x^k(v^k(\xi))| \neq 0, \\ \mathrm{Id}, & |D x^k(v^k(\xi))| = 0. \end{cases}$$

This predictor is total variation preserving, i.e., satisfies the non-strict variant of (3.13). If we have $|(\nabla v^k(\xi))^{-1}||\nabla v^k(\xi)| \leq 1$ for all $\xi \in \Omega$ (e.g., v^k is a simple translation and rotation), this preservation is strict, and the true temporal couplings constructed analogously based on a true displacement field $\bar{v}^k$, satisfy (3.14). Moreover, $\|W_k\|^2, \|T_k\|^2 \leq \Lambda_{\mathcal{V}_k}$ and

$$\left\| \frac{\eta_k}{\eta_{k+1}} K_k - T_k^* K_{k+1} W_k \right\|^2 \leq \tilde{C}_k \|D\|^2$$

for

$$\tilde{C}_k := \max\left(\frac{\eta_k}{\eta_{k+1}}, \sup_{\xi \in \Omega \setminus \Omega_0} \left| \frac{\eta_k}{\eta_{k+1}} - \frac{|\det \nabla(v^k)^{-1}(\xi)| |\nabla v^k((v^k)^{-1}(\xi))^* D x^k(\xi)|}{|D x^k(\xi)|} \right|^2 \right).$$

with $\Omega_0 := \{\xi \in \Omega \mid D x^k(v^k(\xi)) = 0\}$.

Proof. Note that due to $\langle D\bar{x}^k, \bar{y}^k \rangle \leq \|D\bar{x}^k\|_{2,1}\|\bar{y}^k\|_{2,\infty}$ the strict variant of the total variation preservation (3.13) can now be written pointwise

$$(3.16) \quad \alpha|D x^k(\xi)| = \langle D x^k(\xi), y^k(\xi) \rangle \text{ and } |y^k(\xi)| \leq \alpha \text{ for a.e. } \xi \in \Omega$$
$$\implies \alpha|D\check{x}^{k+1}(\xi)| = \langle D\check{x}^{k+1}(\xi), \check{y}^{k+1}(\xi) \rangle \text{ and } |\check{y}^{k+1}(\xi)| \leq \alpha \text{ for a.e. } \xi \in \Omega.$$

Let $\xi \in \Omega$ be an arbitrary point where the antecedent (3.16) holds. When $|Dx^k(v^k(\xi))| = 0$, we deduce that $|D\check{x}^{k+1}(\xi)| = 0$ and so $\alpha|D\check{x}^{k+1}(\xi)| = \langle D\check{x}^{k+1}(\xi), \check{y}^{k+1}(\xi)\rangle$ holds trivially. Suppose then that $|Dx^k(v^k(\xi))| \neq 0$. Since $0 < \alpha|Dx^k(\zeta)| = \langle Dx^k(\zeta), y^k(\zeta)\rangle$ with $|y^k(\zeta)| \leq \alpha$ for all ζ, we must have $y^k(\zeta) = \alpha|Dx^k(\zeta)|^{-1}Dx^k(\zeta)$. Taking $\zeta = v^k(\xi)$, we obtain $y^k(v^k(\xi)) = \alpha|Dx^k(v^k(\xi))|^{-1}Dx^k(v^k(\xi))$. By definition of $\check{x}^k$ and $\check{y}^k$ we obtain, as required

$$\langle D\check{x}^{k+1}(\xi), \check{y}^{k+1}(\xi)\rangle = \langle \nabla v^k(\xi)^* Dx^k(v^k(\xi)), t^k(\xi) y^k(v^k(\xi))\rangle$$

$$= \left\langle Dx^k(v^k(\xi)), \frac{|\nabla v^k(\xi)^* Dx^k(v^k(\xi))|}{|Dx^k(v^k(\xi))|} y^k(v^k(\xi))\right\rangle$$

$$= |\nabla v^k(\xi)^* Dx^k(v^k(\xi))| \left\langle \frac{Dx^k(v^k(\xi))}{|Dx^k(v^k(\xi))|}, \frac{\alpha Dx^k(v^k(\xi))}{|Dx^k(v^k(\xi))|}\right\rangle$$

$$= \alpha|D\check{x}^{k+1}(\xi)|.$$

Thus the fist part of the consequent of (3.16) holds, proving non-strict total variation preservation. If $|(\nabla v^k(\xi))^{-1}||\nabla v^k(\xi)| \leq 1$, we have $|t^k(\xi)| \leq 1$, so $|\check{y}(\xi)^{k+1}| \leq \alpha$ for all ξ, so the second part of the consequent of (3.16) holds, proving strict total variation preservation. The same argumentation applied to analogous true temporal couplings, proves (3.14).

Moreover, with $\Omega_0 = \{\xi \in \Omega \mid Dx^k(v^k(\xi)) = 0\}$,

$$\|T_k\|^2 = \sup_{\|y^k\|=1} \left(\int_{\Omega\setminus\Omega_0} \left|\frac{|\nabla v^k(\xi)^* Dx^k(v^k(\xi))|}{|Dx^k(v^k(\xi))|}(\nabla v^k(\xi))^{-1}y^k(v^k(\xi))\right|^2 d\xi + \int_{\Omega_0}|y^k(v^k(\xi))|^2 d\xi\right)$$

$$\leq \sup_{\|y^k\|=1} \int_\Omega |y^k(v^k(\xi))|^2 d\xi$$

$$= \sup_{\|y^k\|=1} \int_\Omega |y^k(\xi)|^2 |\det \nabla(v^k)^{-1}(\xi)| d\xi \leq \Lambda_{\mathcal{V}_k}.$$

The upper bound for $\|W_k\|$ is obtained by a similar computation. For any $z^{k+1} \in Y_{k+1}$ and $\xi \in \Omega$, we have

$$\int_\Omega \langle z^{k+1}(\xi), t^k(\xi)y^k(v^k(\xi))\rangle d\xi = \int_\Omega \langle t^k(\xi)^* z^{k+1}(\xi), y^k(v^k(\xi))\rangle d\xi$$

$$= \int_\Omega \int \langle t^k((v^k)^{-1}(\xi))^* z^{k+1}((v^k)^{-1}(\xi)), y^k(\xi)\rangle |\det \nabla(v^k)^{-1}(\xi)| d\xi$$

$$= \int_\Omega \langle |\det \nabla(v^k)^{-1}(\xi)| t^k((v^k)^{-1}(\xi))^* z^{k+1}((v^k)^{-1}(\xi)), y^k(\xi)\rangle d\xi$$

for

$$t^k(\xi)^* = \begin{cases} \frac{|\nabla v^k(\xi)^* Dx^k(v^k(\xi))|}{|Dx^k(v^k(\xi))|}((\nabla v^k(\xi))^{-1})^*, & |Dx^k(v^k(\xi))| \neq 0, \\ \mathrm{Id}, & |Dx^k(v^k(\xi))| = 0. \end{cases}$$

Hence

$$\langle z^{k+1}, T_k y^k\rangle = \int_\Omega \langle z^{k+1}(\xi), t^k(\xi)y^k(v^k(\xi))\rangle d\xi = \int_\Omega \langle (T_k^* z^{k+1})(\xi), y^k\rangle d\xi$$

for

$$(T_k^* z^{k+1})(\xi) = |\det \nabla(v^k)^{-1}(\xi)| t^k((v^k)^{-1}(\xi))^* z^{k+1}((v^k)^{-1}(\xi)).$$

Using this expression for T_k^*, for any $\xi \in \Omega \setminus \Omega_0$ we obtain

$$[T_k^* DW_k x^k](\xi) = T_k^*[\xi \mapsto \nabla v^k(\xi)^* Dx^k(v^k(\xi))](\xi)$$

$$= \frac{|\det \nabla(v^k)^{-1}(\xi)||\nabla v^k((v^k)^{-1}(\xi))^* Dx^k(\xi)|}{|Dx^k(\xi)|} Dx^k(\xi),$$

while for any $\xi \in \Omega_0$ we have $[T_k^* DW_k x^k](\xi) = 0$. Thus

$$
\begin{aligned}
\left\| \frac{\eta_k}{\eta_{k+1}} D - T_k^* DW_k \right\|^2 &= \sup_{\|x^k\|=1} \left(\int_{\Omega \setminus \Omega_0} \left| \frac{\eta_k}{\eta_{k+1}} Dx^k(\xi) - [T_k^* DW_k x^k](\xi) \right|^2 d\xi + \int_{\Omega_0} \left| \frac{\eta_k}{\eta_{k+1}} Dx^k(\xi) \right|^2 d\xi \right) \\
&= \sup_{\|x^k\|=1} \left(\int_{\Omega \setminus \Omega_0} \left| \frac{\eta_k}{\eta_{k+1}} - \frac{|\det \nabla (v^k)^{-1}(\xi)| |\nabla v^k((v^k)^{-1}(\xi))^* Dx^k(\xi)|}{|Dx^k(\xi)|} \right|^2 |Dx^k(\xi)|^2 d\xi \right. \\
&\quad \left. + \int_{\Omega_0} \left| \frac{\eta_k}{\eta_{k+1}} \right|^2 \left| Dx^k(\xi) \right|^2 d\xi \right) \\
&\le \tilde{C}_k \sup_{\|x^k\|=1} \int_{\Omega} |Dx^k(\xi)|^2 d\xi = \tilde{C}_k \|D\|^2. \quad \square
\end{aligned}
$$

Remark 3.6. Note that if $\eta_k = \eta_{k+1}$ and $v^k = \mathrm{Id}$ (i.e. there is no prediction), then we can choose $\tilde{C}_k = 1$.

Remark 3.7 (Strict Greedy predictor). During the review process, we realised that a slightly modified version of the dual predictor of Lemma 3.5 strictly preserves total variation. Indeed, take

$$
\breve{y}^{k+1}(\xi) := [T_k y^k](\xi) := \frac{\langle Dx^k(v^k(\xi)), y^k(v^k(\xi)) \rangle}{|Dx^k(v^k(\xi))|} \frac{D\breve{x}^{k+1}(\xi)}{|D\breve{x}^{k+1}(\xi)|},
$$

where we can replace $D\breve{x}^{k+1}(\xi)/|D\breve{x}^{k+1}(\xi)|$ by any unit vector if $D\breve{x}^{k+1}(\xi) = 0$. Then $|\breve{y}^{k+1}(\xi)| \le |y^k(v^k(\xi))| \le \alpha$ and $\langle D\breve{x}^{k+1}(\xi), \breve{y}^{k+1}(\xi) \rangle = \alpha |D\breve{x}^{k+1}(\xi)|$ if $\langle Dx^k(v^k(\xi)), y^k(v^k(\xi)) \rangle = \alpha |Dx^k(v^k(\xi))|$.

We call this the Strict Greedy predictor, as $\breve{y}^{k+1}(\xi)$ has the same direction as $D\breve{x}^{k+1}(\xi)$ with a scaling baed on the total variation attainment discrepancy of the previous step. Although this is a theoretically highly satisfying predictor, we do not include a derivation of the prediction error bounds, as the practical performance is comparable to our other proposed predictors.

POINTWISE PRESERVATION IN L^2 BY ROTATION

An alternative way to preserve total variation following a primal predictor is by implementing a suitable rotation on the dual variable. We use this idea to construct a dual predictor as in the following lemma, which uses the same spaces $X_k \subset L^2(\Omega)$ and $Y_k \subset L^2(\Omega; \mathbb{R}^m)$ as Lemma 3.5.

Lemma 3.8. Let $\mathcal{V}_k \subset H^1(\Omega; \Omega)$ be a set of bijective displacement fields satisfying $(v^k)^{-1} \in H^1(\Omega; \Omega)$ for all $v^k \in \mathcal{V}_k$, and

$$
\Lambda_{\mathcal{V}_k} := \sup_{v^k \in \mathcal{V}_k, \xi \in \Omega} |\det \nabla (v^k)^{-1}(\xi)| < \infty.
$$

Let $x^k \in X_k$, $y^k \in Y_k$, $v^k \in \mathcal{V}_k$, and $(\breve{x}^{k+1}, \breve{y}^{k+1}) = P_k(x^k, y^k) = (W_k x^k, T_k y^k + b_k)$ where the primal predictor $W_k : X_k \to X_{k+1}$ and the affine dual predictor consisting of $T_k : Y_k \to Y_{k+1}$ and $b_k \in Y_{k+1}$, are defined pointwise for $\xi \in \Omega$ by

$$
(3.17) \qquad (W_k x^k)(\xi) = (x^k \circ v^k)(\xi), \qquad (T_k y^k)(\xi) = \begin{cases} R_{\theta_\xi}(y^k(\xi)), & Dx^k(\xi) \ne 0, \\ 0, & Dx^k(\xi) = 0 \end{cases}
$$

and

$$
b_k(\xi) := \begin{cases} \alpha \frac{D\breve{x}^{k+1}(\xi)}{\|D\breve{x}^{k+1}(\xi)\|_2}, & Dx^k(\xi) = 0, D\breve{x}^{k+1}(\xi) \ne 0, \\ 0, & \text{otherwise,} \end{cases}
$$

where R_{θ_ξ} is a rotation operator defined by an oriented angle θ_ξ from $Dx^k(\xi)$ to $D\breve{x}^{k+1}(\xi)$ such that

$$
D\breve{x}^{k+1}(\xi) = c_\xi R_{\theta_\xi}(Dx^k(\xi)) \quad \text{for some} \quad c_\xi \ge 0.
$$

This predictor is strictly total variation preserving, i.e., satisfies the strict variant of (3.13). Moreover, true temporal couplings constructed analogously based on a true displacement field $\bar{v}^k$, satisfy (3.14). Moreover, $\|W_k\|^2 \le \Lambda_{\mathcal{V}_k}, \|T_k\|^2 = 1$, and

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 \le \tilde{C}_k \|D\|^2 \quad for \quad \tilde{C}_k := \sup_{\xi \in \Omega \setminus \Omega_0}\left|\frac{\eta_k}{\eta_{k+1}} - c_\xi,\right|^2$$

where $\Omega_0 := \{\xi \in \Omega \mid Dx^k(\xi) = 0\}$.

Proof. Let $\xi \in \Omega$ satisfy the antecedent of the L^2 characterisation (3.16) of the strict variant of (3.13). If $Dx^k(\xi) = 0$, the consequent of (3.16) is clearly true for ξ by construction. So suppose $Dx^k(\xi) \ne 0$. Since $\check{y}^{k+1}(\xi) = T_k y^k(\xi) = R_{\theta_\xi}(y^k(\xi))$ and $D\check{x}^{k+1}(\xi) = c_\xi R_{\theta_\xi}(Dx^k(\xi))$ for a $c_\xi \ge 0$ such that $|D\check{x}^{k+1}(\xi)| = c_\xi|Dx^k(\xi)|$, then, as required by (3.16),

$$\begin{aligned}
\langle D\check{x}^{k+1}(\xi), \check{y}^{k+1}(\xi)\rangle &= \langle c_\xi R_{\theta_\xi}(Dx^k(\xi)), R_{\theta_\xi}(y^k(\xi))\rangle \\
&= c_\xi \langle Dx^k(\xi), y^k(\xi)\rangle = c_\xi \alpha|Dx^k(\xi)| = \alpha|D\check{x}^{k+1}(\xi)|.
\end{aligned}$$

Clearly $|\check{y}(\xi)^{k+1}| \le \alpha$, so the consequent of (3.16) holds for ξ, as required. The same argumentation applied to analogously constructed true temporal couplings, proves (3.14).

Computing for upper bounds of $\|W_k\|^2$ and $\|T_k\|^2$ is straightforward. Moreover, by the definition of predictors, for any $\xi \in \Omega$ with $Dx^k(\xi) \ne 0$, we have

$$\begin{aligned}
T_k^* DW_k x^k(\xi) = T_k^* D\check{x}^{k+1}(\xi) &= T_k^*(c_\xi R_{\theta_\xi}(Dx^k(\xi)) \\
&= R_{-\theta_\xi}(c_\xi R_{\theta_\xi}(Dx^k(\xi)) = c_\xi Dx^k(\xi).
\end{aligned}$$

On the other hand, if $Dx^k(\xi) = 0$, by the construction of T_k^*, clearly $T_k^* DW_k x^k(\xi) = 0$. Thus

$$\begin{aligned}
\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 &= \sup_{\|x^k\|=1}\int_\Omega \left|\frac{\eta_k}{\eta_{k+1}}Dx^k(\xi) - T_k^* DW_k x^k(\xi)\right|^2 d\xi \\
&= \sup_{\|x^k\|=1}\int_{\Omega \setminus \Omega_0} \left|\frac{\eta_k}{\eta_{k+1}}Dx^k(\xi) - c_\xi Dx^k(\xi)\right|^2 d\xi \le \tilde{C}_k\|D\|^2. \quad \square
\end{aligned}$$

Remark 3.9. If $\eta_k = \eta_{k+1}$ and $c_\xi = 1$ (i.e. $\|D\check{x}^{k+1}(\xi)\| = \|Dx^k(\xi)\|$) for all ξ, then we can choose $\tilde{C}_k = 0$.

GLOBAL PRESERVATION FOR GENERAL OPERATORS

We now study a primal-dual predictor that preserves total variation (defined with a general operator D) vectorwise. To do so, we need a left-invertible modification $\check{D}$ of the (discretised, differential) operator D. We let X_k be a Hilbert space, $Y_k \subset L^2(\Omega; \mathbb{R}^m)$, and $K_k = D \in \mathbb{L}(X_k; Y_k)$ for all k for an arbitrary domain Ω, such that the $2, 1$-norm is defined.

Lemma 3.10. Let $x^k \in X_k, y^k \in Y_k, W_k \in \mathbb{L}(X_k; X_{k+1}), D, \check{D} \in \mathbb{L}(X_k; Y_k)$. Suppose $(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k)$ where, for $Q_k := \|\check{D}\check{x}^{k+1}\|_{2,1}\|Dx^k\|_{2,1}^{-1}\mathrm{Id} \in \mathbb{L}(Y_k; Y_k)$, the primal-dual predictor

$$(3.18) \qquad P_k(x^k, y^k) = (W_k x^k, \check{D}z^k) \text{ for a } z^k \in X_k \text{ satisfying } W_k^* \check{D}^* \check{D}z^k = D^* Q_k y^k$$

This predictor is total variation preserving, i.e., satisfies (3.13). If W_k is invertible and $\operatorname{ran}\check{D}^* = \operatorname{ran}(W_k^*)^{-1}D^*$, then we can write (3.18) in explicit form as

$$P_k(x^k, y^k) = (W_k x^k, T_k y^k) \text{ where } T_k = \check{D}(\check{D}^*\check{D})^{-1}(W_k^*)^{-1}D^* Q_k.$$

Furthermore,

$$\|T_k\|^2 \leq \frac{\|\check{D}(\check{D}^*\check{D})^{-1}(W_k^*)^{-1}D^*\|\|\check{D}\check{x}^{k+1}\|_{2,1}}{\|Dx^k\|_{2,1}} \quad \text{and}$$

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^*K_{k+1}W_k\right\|^2 \leq \tilde{C}_k := \left\|\frac{\eta_k}{\eta_{k+1}}D - Q_k D W_k^{-1}(\check{D}^*\check{D})^{-1}\check{D}^*DW_k\right\|^2.$$

Proof. Using the definition of predictors, we have

$$\langle \check{D}\check{x}^{k+1}, \check{y}^{k+1}\rangle = \langle \check{D}W_k x^k, \check{D}z^k\rangle = \langle x^k, W_k^*\check{D}^*\check{D}z^k\rangle$$

$$= \langle x^k, D^*Q_k y^k\rangle = \langle Dx^k, Q_k y^k\rangle = \frac{\|\check{D}\check{x}^{k+1}\|_{2,1}}{\|Dx^k\|_{2,1}}\langle Dx^k, y^k\rangle = \alpha\|\check{D}\check{x}^{k+1}\|_{2,1},$$

where in the last equality we used $\alpha\|Dx^k\|_{2,1} = \langle Dx^k, y^k\rangle$. Now, let W_k be invertible. Then using ran $\check{D}^* = \mathrm{ran}(W_k^*)^{-1}D^*$ and (3.18) we have

$$W_k^*\check{D}^*\check{D}z^k = D^*Q_k y^k \implies T_k = \check{D}(\check{D}^*\check{D})^{-1}(W_k^*)^{-1}D^*Q_k.$$

With this, the bounds for $\|T_k\|^2$ and $\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^*K_{k+1}W_k\|^2$ follow immediately. $\square$

> **Example 3.11.** For total variation, it is natural to take D with (discrete) Neumann boundary conditions. This operator is not left-invertible (ran $D^* \neq X_k$): constant functions are in the kernel. To form a left-invertible $\check{D}$, for example, with forward differences discretisation, it suffices to take it with (discrete) Dirichlet boundary conditions. If constant functions are an invariant subspace of $(W_k^*)^{-1}$, then also D with forward differences discretisation and (discrete) Neumann boundary conditions satisfies the invertibility condition.

Remark 3.12. If $D = \check{D}$, then

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^*K_{k+1}W_k\right\|^2 = \left\|\frac{\eta_k}{\eta_{k+1}}D - Q_k D\right\|^2 = \left|\frac{\eta_k}{\eta_{k+1}} - \frac{\|\check{D}\check{x}^{k+1}\|_{2,1}}{\|Dx^k\|_{2,1}}\right|^2\|D\|^2.$$

Thus, if we further set $\eta_k = \eta_{k+1}$ and $W_k = \mathrm{Id}$ (i.e., there are no predictions), then we can take $\tilde{C}_k = 0$.

3.3 INNER PRODUCT PRESERVING PREDICTORS

Instead of preserving the total variation after each prediction, we can impose that only the angles are preserved, that is,

$$\langle D\check{x}^{k+1}, \check{y}^{k+1}\rangle = \langle Dx^k, y^k\rangle \tag{3.19}$$

for all k. We refer to such predictors as *inner product preserving*. This is motivated by the fact that, in the static case, the inner product $\langle Dx^{k+1}, y^{k+1}\rangle$ converges to the total variation. Thus, if we are given a primal predictor, we mitigate the error caused by the predictions by computing for a dual prediction such that equality in (3.19) is preserved.

We start by noting that the predictors from Section 3.2 preserve inner products by a simple rescaling of the dual predictors:

- The predictors of Lemma 3.5 with $t^k(\xi) = |\det \nabla v^k(v^k(\xi))|(\nabla v^k(\xi))^{-1}$.

- The rotating predictors of Lemma 3.8 with $(T_k y^k)(\xi) = c_\xi^{-1}R_{\theta_\xi}(y^k(\xi))$.

- The abstract predictors of Lemma 3.10 with $Q_k = \mathrm{Id}$.

We omit the proofs and bounds, as we consider inner product preservation a weaker result than total variation preservation.

(GREEDY) COMPONENT-WISE PRESERVATION IN FINITE DIMENSIONS

An alternative approach to guarantee the preservation of the inner product is a component-wise update of the dual variable. This strategy directly enforces the point-wise equality $\langle D\check{x}^{k+1}(\xi), \check{y}^{k+1}(\xi)\rangle = \langle Dx^k(\xi), y^k(\xi)\rangle$ down to each individual component of ξ. Although straightforward, it does not consider the potential impact on other components, hence the prediction is greedy. However, as will be shown later, this method can still exhibit good numerical performance in practice.

Lemma 3.13. *Let $x^k \in \mathbb{R}^n$, $y^k \in \mathbb{R}^m$ and $D \in \mathbb{R}^{m\times n}$. For $W_k \in \mathbb{R}^{n\times n}$, if $(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k) = (W_k x^k, T_k y^k)$ where the primal prediction $\check{x}^{k+1} = W_k x^k$ and, for each component i, the dual prediction*

$$(3.20) \qquad (T_k y^k)_i = \begin{cases} \frac{(Dx^k)_i}{(D\check{x}^{k+1})_i} y_i^k, & |(D\check{x}^{k+1})_i| > \varepsilon \\ y_i^k, & |(D\check{x}^{k+1})_i| \le \varepsilon \end{cases},$$

for some tolerance $\varepsilon > 0$, then $(D\check{x}^{k+1})_i \check{y}_i^{k+1} = (Dx^k)_i y_i^k$ whenever $|(D\check{x}^{k+1})_i| > \varepsilon$. Moreover, if $(D\check{x}^{k+1})_i \ne 0$ for all i, then

$$\|T_k\|^2 \le \max_i \frac{|(Dx^k)_i|^2}{|(D\check{x}^{k+1})_i|^2}$$

and

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 \le \tilde{C}_k\|D\|^2 \quad for \quad \tilde{C}_k := \left|\frac{\eta_k}{\eta_{k+1}} - 1\right|^2.$$

Proof. It is straightforward to show that $(D\check{x}^{k+1})_i \check{y}_i^{k+1} = (Dx^k)_i y_i^k$ whenever $(D\check{x}^{k+1})_i \ne 0$. Now, if $(D\check{x}^{k+1})_i \ne 0$ for all i, then

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 = \sup_{\|x^k\|=1}\left\|\frac{\eta_k}{\eta_{k+1}}Dx^k - Dx^k\right\|^2 \le \tilde{C}_k\|D\|^2. \qquad \square$$

The computation for the upper bound of $\|T_k\|^2$ is straightforward.

Remark 3.14. By simply setting $\eta_k = \eta_{k+1}$, we can choose $\tilde{C}_k = 0$.

3.4 DUAL SCALING PREDICTORS

We now consider another type of predictor that uses pointwise scaling in the dual variable. For these predictors, we do not have any explicit preservation results, although, as we shall see, they perform numerically remarkably well.

Lemma 3.15. *For a domain $\Omega \subset \mathbb{R}^n$, consider the subspaces $X_k \subset L^2(\Omega)$ and $Y_k \subset L^2(\Omega; \mathbb{R}^n)$ equipped with the L^2-norm. Let $x^k \in X_k$, $y^k \in Y_k$, $K_k = D \in \mathbb{L}(X_k; Y_k)$ a differential operator, and $(\check{x}^{k+1}, \check{y}^{k+1}) = P_k(x^k, y^k) = (W_k x^k, T_k y^k)$ for any primal predictor $W_k \in \mathbb{L}(X_k; X_{k+1})$, and dual predictor $T_k \in \mathbb{L}(Y_k; Y_{k+1})$, defined pointwise by*

$$(3.21) \qquad (T_k y^k)(\xi) = c_k(\xi) y^k(\xi).$$

where $c_k \in L^\infty(\Omega)$. Then

$$\|T_k\|^2 = \bar{C}_k \quad and \quad \left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 \le 2\left(\tilde{C}_k + \bar{C}_k\|\operatorname{Id} - W_k\|^2\right)\|D\|^2$$

for

$$\tilde{C}_k := \sup_{\xi\in\Omega}\left|\frac{\eta_k}{\eta_{k+1}} - c_k(\xi)\right|^2 \quad and \quad \bar{C}_k := \sup_{\xi\in\Omega}|c_k(\xi)|^2.$$

Proof. It is straightforward to verify that $\|T_k\|^2 = \bar{C}_k$. Now,

$$\left\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\right\|^2 = \sup_{\|x^k\|=1} \int_\Omega \left|\frac{\eta_k}{\eta_{k+1}}Dx^k(\xi) - c_k(\xi)DW_k x^k(\xi)\right|^2 d\xi$$

$$\leq 2 \sup_{\|x^k\|=1} \int_\Omega \left(\left|\frac{\eta_k}{\eta_{k+1}} - c_k(\xi)\right|^2 |Dx^k(\xi)|^2 + |c_k(\xi)|^2 |D(\mathrm{Id} - W_k)x^k(\xi)|^2\right) d\xi$$

$$\leq 2\left(\tilde{C}_k + \bar{C}_k \|\mathrm{Id} - W_k\|^2\right)\|D\|^2. \quad \square$$

Remark 3.16. If $\eta_k = \eta_{k+1}$, and $|c_k(\xi)| \to 1$ for all ξ as $k \to \infty$ whenever $W_k \to \mathrm{Id}$, then $\|\frac{\eta_k}{\eta_{k+1}}K_k - T_k^* K_{k+1} W_k\| \to 0$.

In Section 4, we test in application a dual scaling predictor tailored in such a way that its scaling factors invalidate the dual at points where there are significant changes in the primal.

4 NUMERICAL EXPERIMENTS

We conducted experiments with our method in two applications: the image stabilisation (compare with [33]) and the dynamic Positron Emission Tomography (PET) reconstruction.

In our image stabilisation experiment, we sequentially process highly noisy randomly displaced sub-images of a bigger image. For simplicity we assume to have access to noisy measurements of the displacements. Leveraging these displacements in the primal predictors, our objective is to achieve real-time denoising of the displaced sub-images with total variation regularisation. In dynamic PET imaging, our focus shifts to an unknown density (image) undergoing rotational motion during the measurement process. Our task involves reconstructing this density using data obtained from its partial Radon transform, again subject to Poisson noise. We also assume access to noisy measurements of the rotation angle. If measurements of the displacement or the rotation is not otherwise available, they can be incorporated into our model following the optical flow displacement estimation approach in [33], for PET by adapting patient body motion tracking such as that in [17].

4.1 PREDICTORS

For image stabilisation, the primal prediction is simply $\check{x}^{k+1} := x^k \circ v^k$, where v^k represents the noisy measurement of an unknown displacement $\bar{v}^k$ of the subimage for frame k. For the dynamic PET experiment, the primal prediction is $\check{x}^{k+1}(\xi) := R^k_{\theta_\xi}(x^k(\xi))$, where $R^k_{\theta_\xi}$ models the random rotational motion. Based on the theoretical framework outlined in Section 3, we evaluate several choices for the dual predictors:

- *Rotation*: A total variation preserving predictor that employs rotation on the dual variable as described in Lemma 3.8.

- *Greedy*: An inner product preserving predictor by components as described in Lemma 3.13.

- *Strict Greedy*: A strict total variation preserving predictor as described in Remark 3.7.

- *Dual Scaling*: A predictor enforces component-wise scaling as described in Lemma 3.15 with

$$c_k(\xi) := 1 - \chi_k \nu_k(x_\delta^{k+1}(\xi)) \quad \text{where} \quad x_\delta^{k+1}(\xi) := \frac{|\check{x}^{k+1}(\xi) - x^k(\xi)|}{\max\{10^{-12}, \max_\xi |\check{x}^{k+1}(\xi) - x^k(\xi)|\}},$$

for a scalar χ_k and an "activation" function ν_k satisfying $\nu_k(0) = 0$ and $\nu_k(1) = 1$. For experiments where the test image or phantom has mostly flat regions (e.g., Shepp-Logan phantom below),

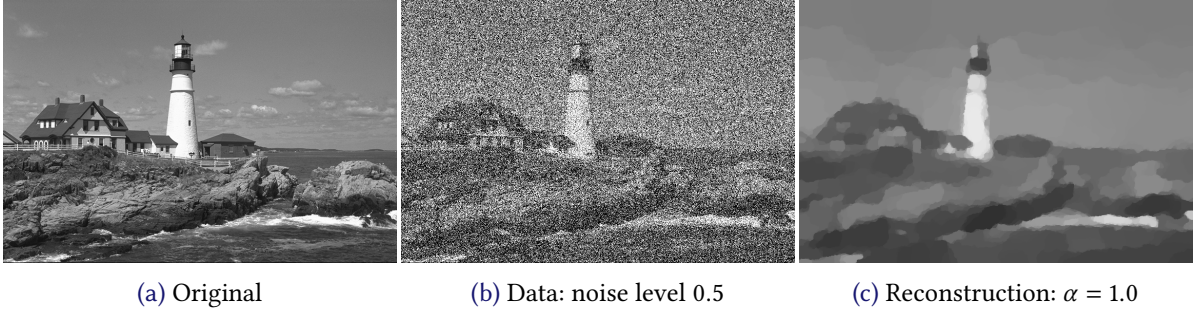(a) Original   (b) Data: noise level 0.5   (c) Reconstruction: $\alpha = 1.0$

Figure 1: Test image, added noise, and stationary reconstruction for comparison.

we use $\chi_k = 1.0$ and $v_k(\cdot) = (1 + e^{-1000(\cdot - 0.05)})^{-1}$. Otherwise (e.g., for lighthouse test image and brain phantom), we fix $\chi_k = 0.75$ and $v_k(\cdot) = 1 - |\cdot - 1|^{1/5}$.

The idea is that the dual predictor attempts to move $\|\check{y}^k(\xi_k)\|$ towards either α or 0 when the primal predictor has changed the corresponding primal variable. However, when the changes in the primal variable become small, the activation function also causes the changes in the dual variable to become small, allowing it to stabilise, and the prediction errors to become small via $|c_k(\xi)| \to 1$ as in Remark 3.16.

We compare these proposed predictors to

- *No Prediction*, i.e., both primal and dual predictors are identity maps;

- *Primal Only* predictor with identity dual prediction;

- *Zero Dual* predictor, i.e., $\check{y}^k = 0$; and the

- *Proximal* (old) dual predictor from [33].

For the latter, we use $\tilde{G}_k^* = G_k^* + \frac{\tilde{\rho}_k}{2}\|\cdot\|_{L^2(\Omega)}^2$ in the proximal prediction step, and a "phantom" $\rho = 100$, as discussed in [33], to allow for larger dual step lengths. The Zero Dual predictor is motivated by the staircasing effect of total variation: it promotes flat regions, where the dual variable can be close to zero. As we will see, it performs remarkably well when there is movement in the images, but fails to stabilise when the movement is stopped.

4.2 IMAGE STABILISATION

For image stabilisation, we assume to be given in each frame a noisy measurement $z_k \in X_k$ of a true image $\bar{z}_k \in X_k$ and a noisy measurement $v^k \in V_k$ of a true displacement field $\bar{v}^k \in V_k$. We require the measured displacement fields v^k to be bijective. The finite-dimensional subspaces $X_k \subset L^2(\Omega)$, $Y_k \subset (\Omega; \mathbb{R}^2)$, and $V_k \subset L^2(\Omega; \Omega) \cap C^2(\Omega; \Omega)$ on a domain $\Omega \subset \mathbb{R}^2$ we equip with the L^2-norm. We set the objective functions J_k in (1.1) by taking

$$F_k(x) := \frac{1}{2}\|x - z_k\|^2, \quad E_k(x) := 0, \quad \text{and} \quad (G_k \circ K_k)(x) := \alpha\|D_k x\|_{2,1}$$

where D_k is a discretised differential operator, and α is the regularisation parameter for total variation.

| Predictor | Average PSNR | | PSNR | Average SSIM | | SSIM |
	iter 1	iter 500	95% CI	iter 1	iter 500	95% CI
Dual Scaling	**22.6959**	**27.9238**	**27.0909–28.7567**	**0.6697**	**0.8101**	**0.7973–0.8229**
Greedy	21.7029	26.5375	25.8964–27.1786	0.6509	0.7877	0.7740–0.8014
No Prediction	19.9162	24.2983	23.3691–25.2275	0.6201	0.7629	0.7436–0.7822
Primal Only	21.7029	26.5374	25.8963–27.1785	0.6509	0.7877	0.7740–0.8014
Proximal	21.5815	26.3912	25.7342–27.0482	0.6537	0.7955	0.7808–0.8102
Rotation	21.8185	26.6875	26.0523–27.3227	0.6588	0.7963	0.7834–0.8092
Strict Greedy	21.6471	26.4771	25.8202–27.1340	0.6572	0.7989	0.7844–0.8134
Zero Dual	21.9269	26.8247	26.2399–27.4095	0.5940	0.7012	0.6766–0.7258

Table 1: Average PSNR and SSIM for computational image stabilisation, from the indicated iteration until 10000 iterations. The confidence intervals (CI) are computed starting from the 500th iteration.

| Predictor | Average PSNR | | PSNR | Average SSIM | | SSIM |
	iter 1	iter 500	95% CI	iter 1	iter 500	95% CI
Dual Scaling	**14.7561**	**19.8815**	**19.2024–20.5606**	0.5883	**0.8309**	**0.8102–0.8516**
Greedy	14.7223	19.8249	19.1398–20.5100	0.5882	0.8285	0.8074–0.8496
No Prediction	12.9669	18.0957	17.0723–19.1191	0.4391	0.7106	0.6606–0.7606
Primal Only	14.7340	19.8448	19.1615–20.5281	0.5875	0.8286	0.8075–0.8497
Proximal	14.7176	19.8137	19.1259–20.5015	**0.5894**	0.8278	0.8065–0.8491
Rotation	14.7403	19.8539	19.1730–20.5348	0.5875	0.8291	0.8081–0.8501
Strict Greedy	14.7175	19.8143	19.126–20.5026	0.5883	0.8262	0.8053–0.8471
Zero Dual	14.5527	19.4534	18.9525–19.9543	0.5681	0.7888	0.7814–0.7962

Table 2: Average PSNR and SSIM for dynamic PET reconstruction with Shepp-Logan phantom. from the indicated iteration until 4000 iterations. The confidence intervals (CI) are computed starting from the 500th iteration.

| Predictor | Average PSNR | | PSNR | Average SSIM | | SSIM |
	iter 1	iter 500	95% CI	iter 1	iter 500	95% CI
Dual Scaling	**15.9222**	**21.1804**	**20.8011–21.5597**	**0.4720**	**0.6542**	**0.6353–0.6731**
Greedy	15.8040	20.9289	20.6396–21.2182	0.4677	0.6447	0.6292–0.6602
No Prediction	14.2962	19.7518	19.1786–20.3250	0.4096	0.6021	0.5770–0.6272
Primal Only	15.7985	20.9236	20.6335–21.2137	0.4672	0.6443	0.6287–0.6599
Proximal	15.7758	20.8758	20.5797–21.1719	0.4667	0.6431	0.6274–0.6588
Rotation	15.8037	20.9331	20.6444–21.2218	0.4673	0.6446	0.6291–0.6601
Strict Greedy	15.7772	20.8803	20.5828–21.1778	0.4666	0.6432	0.6273–0.6591
Zero Dual	15.8139	20.9511	20.6682–21.2340	0.4639	0.6372	0.6250–0.6494

Table 3: Average PSNR and SSIM for dynamic PET reconstruction with brain phantom, from the indicated iteration until 4000 iterations. The confidence intervals (CI) are computed starting from the 500th iteration.
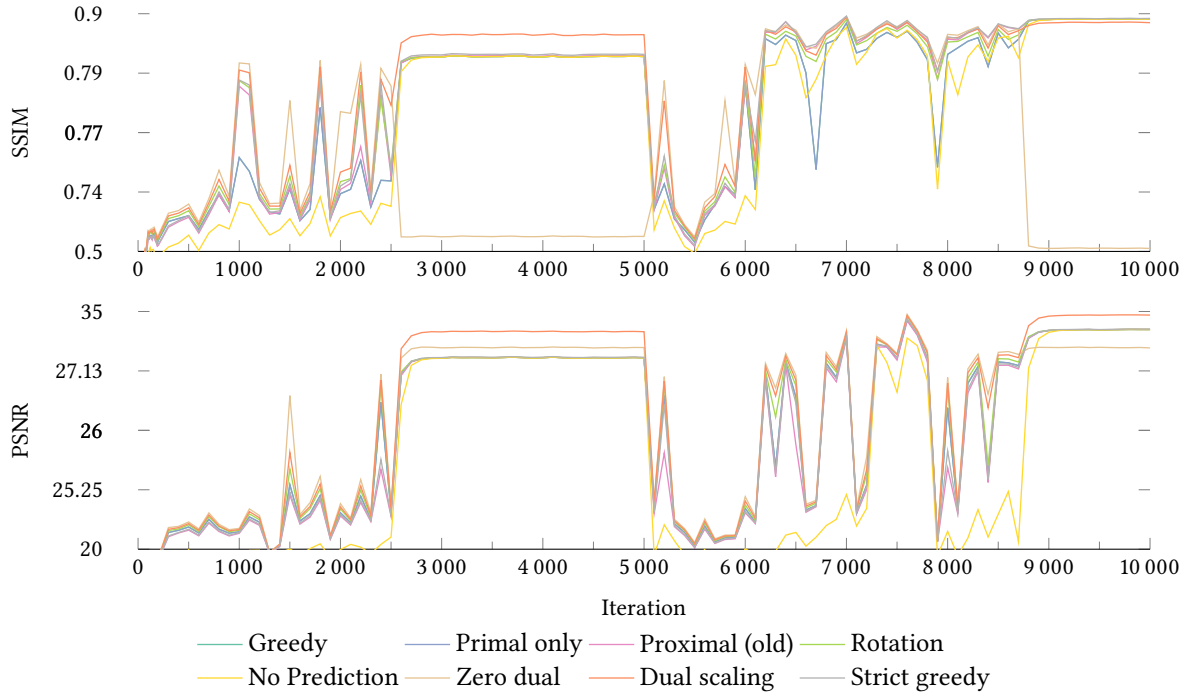
Figure 2: Iteration-wise SSIM and PSNR for the image stabilisation experiment.

NUMERICAL SETUP AND RESULTS

To evaluate our approach, we use as test image the lighthouse image (from the free Kodak image suite [11]) displayed in Figure 1 along with a noisy version and a single-frame total variation reconstruction for comparison. The original size is 768×512 pixels. For our experiments, we pick a 300×200 sub-image moving according to Brownian motion of standard deviation 2. This motion is stopped on two subintervals (frames 2500–5000 and 8700–10000). Thus the displacement fields $\bar{v}^k(\xi) = \xi + d^k$ with $d^k \in \mathbb{R}^2$ are constant in space. We added 50% Gaussian noise (standard deviation 0.5 with original intensities in $[0, 1]$) to the sub-image. To construct the measured displacements available to the algorithm, we add 2.5% Gaussian noise to the true displacements. It is worth mentioning that even on those intervals where the motion is stopped, the displacements available to the algorithm are still not necessarily zero due to introduced noise.

We take the regularisation parameter $\alpha = 0.25$. We set the following parameters for Algorithm 1:
  – Step length parameter $\tau = 0.01$, as well as $\Lambda = \Theta = 1$.
  – Primal strong convexity factor $\gamma = 1$ and, generally, dual factor $\rho = 0$.
  – Maximal σ and estimate $\|K_k\| \leq \sqrt{8}$ for forward-differences discretisation of $K_k = D$ with cell width $h = 1$.

We always take zero as the initial iterate (primal and dual).

We implemented our algorithms in Julia, and performed our experiments on a mid-2022 MacBook Air with 16GB RAM and eight CPU cores. The implementation is available on Zenodo [34].

Figure 3 shows the reconstructed lighthouse subimages. The initial frames are darker and noisier due to the algorithm requiring a certain number of steps to stabilise. The reconstructions without any predictor (a) exhibit the poorest quality. The lighthouse and background features appear blurry and lack definition. The reconstruction generated by the Proximal (old) predictor of [31] (b) appears visually distinct from the others. While it retains some level of detail, the overall image appears smoother and flatter. The reconstructions obtained using the Zero Dual (c) and the Dual Scaling predictors (d) demonstrate improved image quality in intervals where the object is not stable. These methods

(a) No prediction   (b) Proximal (old)   (c) Zero dual   (d) Dual scaling
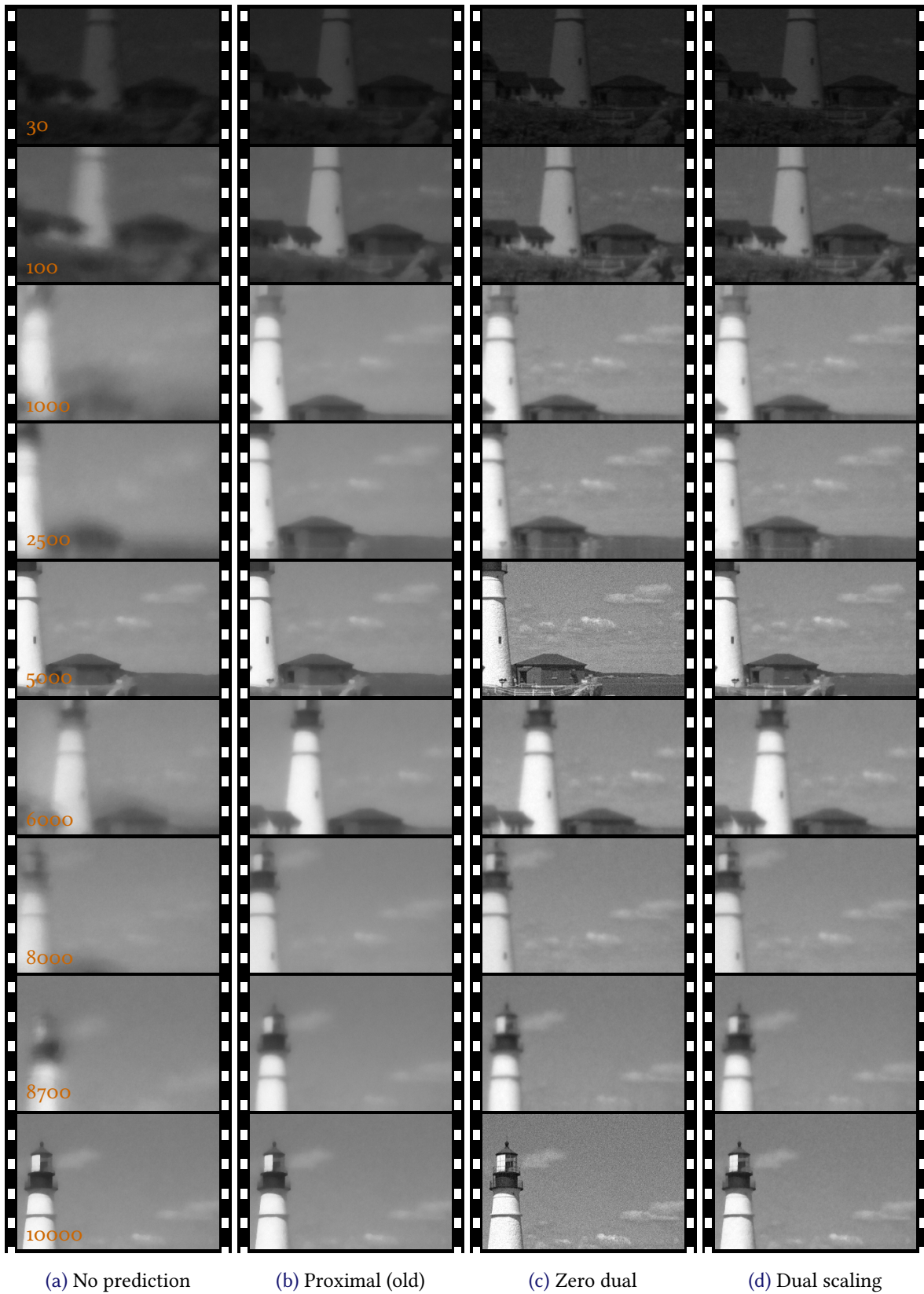
Figure 3: Image stabilisation results for several predictors when $\alpha = 0.25$. The numbers on the left indicate the iteration/frame.

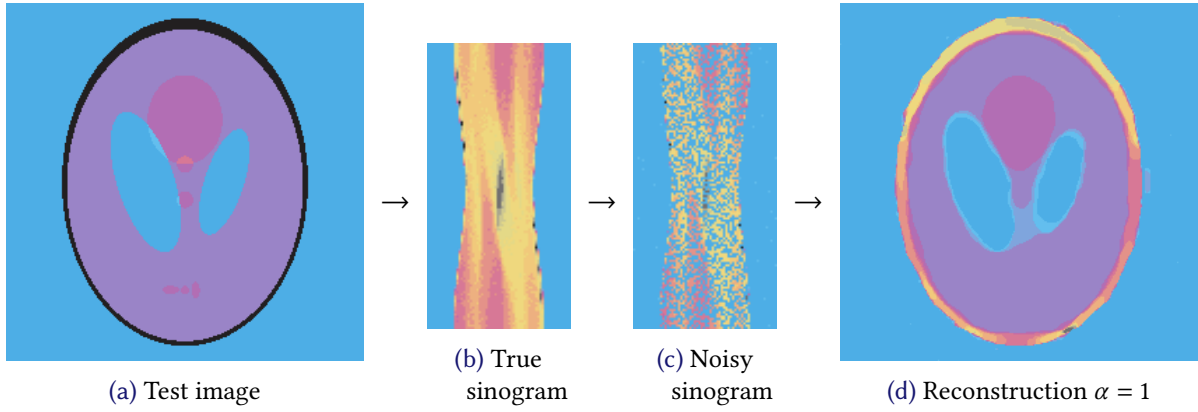(a) Test image   (b) True   (c) Noisy   (d) Reconstruction $\alpha = 1$
                       sinogram   sinogram

Figure 4: Shepp-Logan phantom, true sinogram, noisy subsampled sinogram, and static reconstruction. The colours represent values in $[0, 1]$ as �.

recover sharper edges and finer details, making the lighthouse structure and the clouds more prominent. Howeve, the Zero Dual still gives noisy reconstructions at intervals where the movement is stopped, compared to that of the Dual Scaling predictors.

To quantify these observations, Figure 2 shows the plots of Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for each reconstructed frame. The Proximal (old) dual predictor exhibits lower SSIM and PSNR values compared to those of our proposed Dual Scaling, Greedy, Strict Greedy, and Rotation dual predictors. Moreover, the reconstructions produced by the Dual Scaling predictor attained the highest SSIM and PSNR scores, even during intervals where motion is stopped, surpassing those of the No Prediction method. During stabilisation, even though the true displacement fields are zeroed out, the displacement fields available to the algorithms are still subject to small noise, and the Dual Scaling predictor adapted well even with these small errors. Table 1 additionally shows average, SSIM and PSNR, as well as 95% confidence intervals (CI) for them. The Dual Scaling predictor consistently delivers high image quality scores. The Zero Dual, however, delivers poor SSIM scores, especially on the intervals where the movement is stopped.

In summary, both visual inspection and quantitative metrics demonstrate the effectiveness of our proposed simplified Algorithm 1 and several proposed predictors, compared to the original algorithm from [33], and to employing no prediction at all. This suggests that carefully designed dual predictors can play a crucial role in improving visual quality of the reconstructions.

4.3 DYNAMIC POSITRON EMISSION TOMOGRAPHY

In dynamic PET, we assume to be given in each frame a noisy PET dataset z_k in the data space obtained from a test image $\bar{x}^k \in X_k$. We let assume to be given a noisy θ to form $R_\theta^k : X_k \to X_{k+1}$ that models the noisy rotational motion (by an angle θ about a perturbed center) of pixels between frames. The finite-dimensional subspaces $X_k \subset L^2(\Omega)$, $Y_k \subset (\Omega; \mathbb{R}^2)$, and $V \subset L^2(\Omega; \Omega) \cap C^2(\Omega; \Omega)$ on a domain $\Omega \subset \mathbb{R}^2$ equipped with the L^2-norm. We set the objective functions J_k in (1.1) by taking

$$F_k(x) := \delta_{\mathbb{R}^+}, \quad E_k(x) := \sum_{i=1}^{n} \Big([A_k x^k]_i - [z_k]_i \log([A_k x^k + c_k]_i) \Big), \quad \text{and} \quad (G_k \circ K_k)(x) := \alpha \|D_k x\|_{2,1},$$

where A_k is the forward model based on a partial Radon transform and c_k is a known vector with non-negative entries that models the expected number of background events.
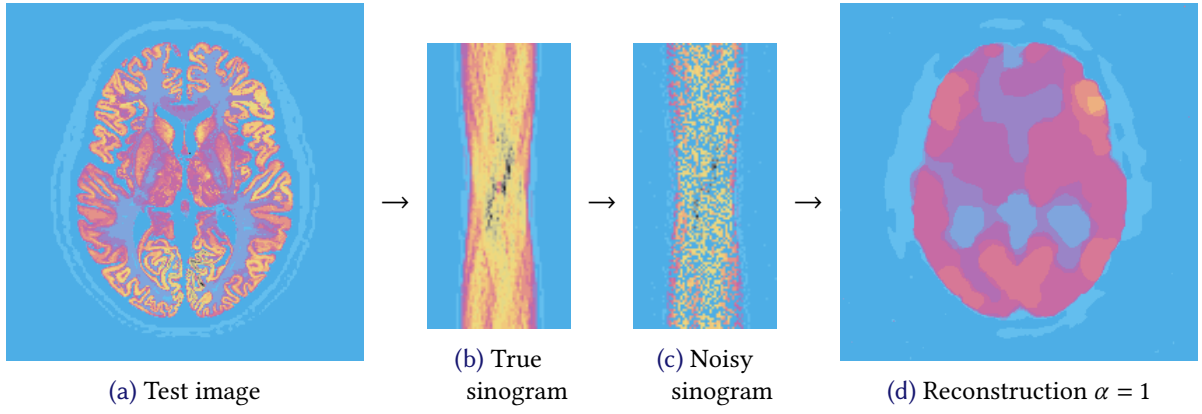
(a) Test image   (b) True sinogram   (c) Noisy sinogram   (d) Reconstruction $\alpha = 1$

Figure 5: Brain phantom [2], true sinogram, noisy subsampled sinogram, and static reconstruction. The colours represent values in $[0, 1]$ as ▬.
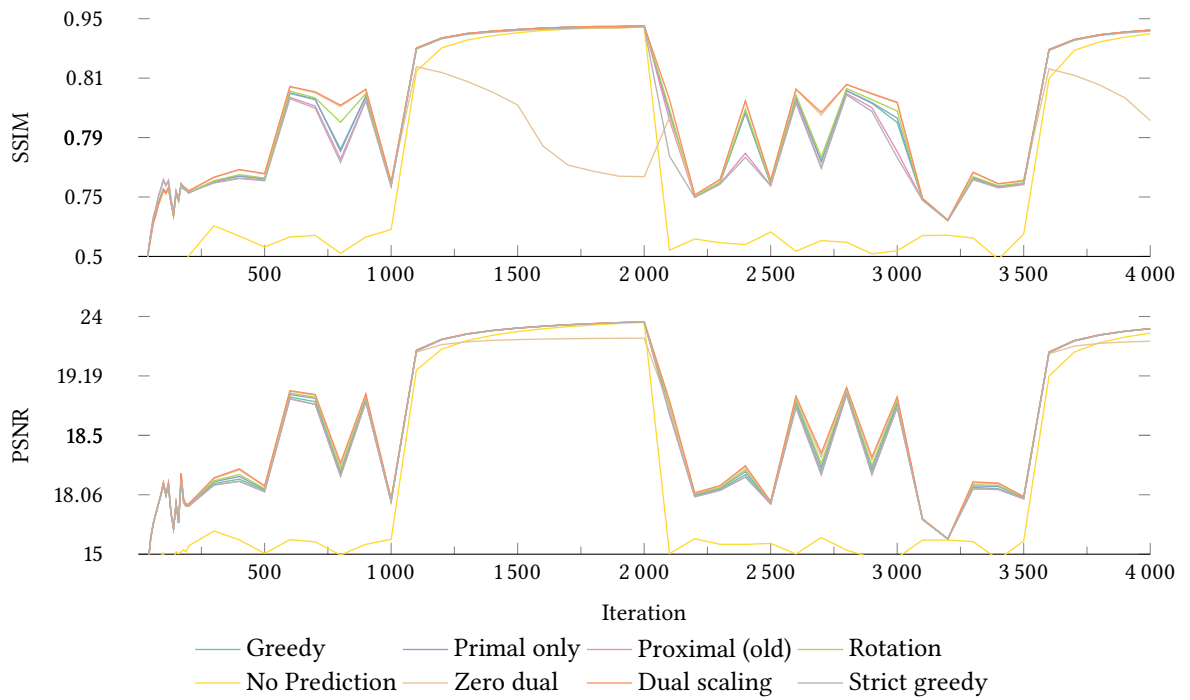

Figure 6: Iteration-wise SSIM and PSNR for PET with Shepp-Logan phantom.

NUMERICAL SETUP AND RESULTS

In our experiments, we use the Shepp-Logan phantom and the high-resolution brain phantom of [2] to generate true and noisy PET datasets, examples of which are displayed in Figure 4. We set both phantoms to have a resolution of 256×256 pixels, while the PET datasets are organised into a sinogram of resolution 128×64. To simulate the random nature of positron emissions, hence PET measurements, we further randomly subsample the sinogram down to 50%. To simulate motion, we rotate the phantom around a randomly chosen axis that shifts from the phantom's centre by a standard deviation of 1. The rotation angles themselves are chosen from a separate Gaussian distribution with a standard deviation of 0.15 radians. We introduce Poisson noise with mean parameter of 0.5 to the PET dataset. To construct the simulated displacement measurements, that are available to the algorithms, we introduced Gaussian noise with a standard deviation of 0.035 radians to the rotation angles and Gaussian noise with a
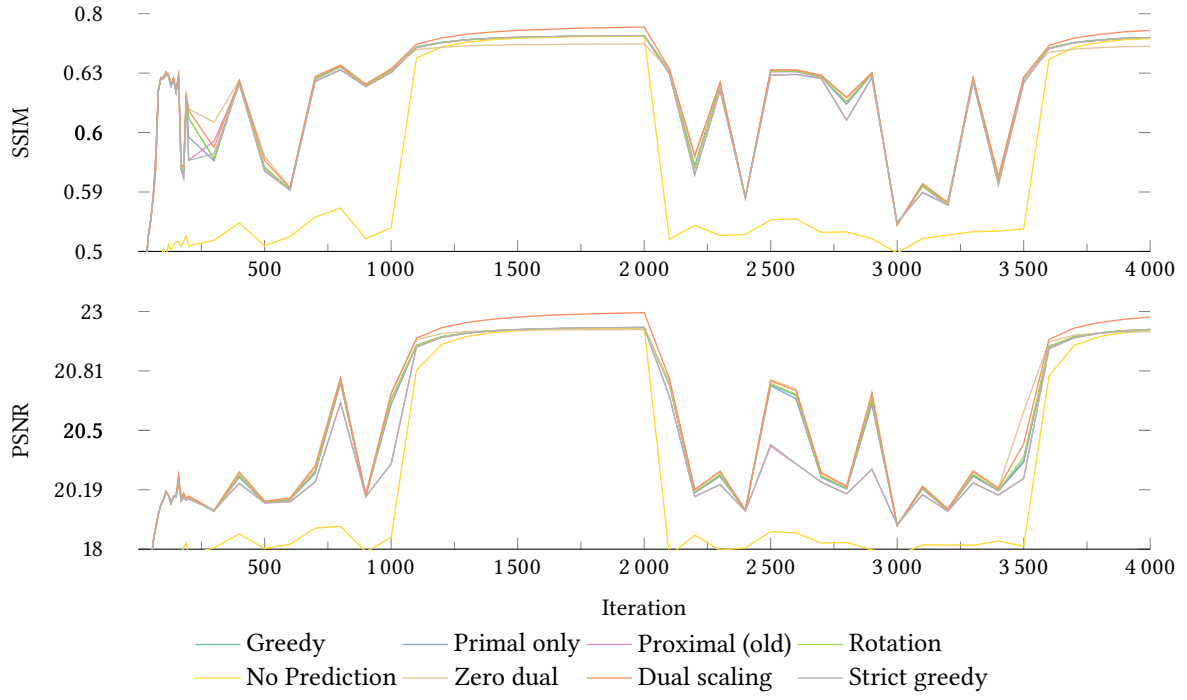
Figure 7: Iteration-wise SSIM and PSNR for PET with brain phantom.

standard deviation of 0.25 to the centre of rotation. The movement is stopped on two subintervals (frames 1000–2000 and 3500–4000). We note again that even in these intervals, the displacements made available to the algorithm are still not necessarily zero due to introduced noise.

Taking the regularisation parameter $\alpha = 0.25$, for Algorithm 1 we take:

- Step length parameter $\tau = 0.003$, as well as $\Lambda = \Theta = \kappa = 1$.
- Primal strong convexity factor $\gamma = 1$ and generally the dual factor $\rho = 0$.
- Maximal σ with the estimate $\|K_k\| \leq \sqrt{8}$ when $K_k = D$ is the forward differences operator with cell width $h = 1$ [7].
- We also fix $L = 300$, experimentally determined to be an upper bound over all iterations k for

$$L_{k+1} = \max \left\{ L_k, 0.9 \frac{\|\nabla E_k(x^k) - \nabla E_k(\check{x}^k)\|}{\|x^k - \check{x}^k\|} \right\}.$$

Figure 8 shows the PET reconstructions of the Shepp-Logan phantom. The initial frames are omitted as they exhibit noisy appearance due to the algorithm requiring a certain number of steps to stabilise. The performance without any predictor (a), is clearly the poorest, producing blurry artefacts around the moving parts. The differences between the other reconstructions are less noticeable. However, a closer inspection reveals that the oval shapes appear more distinct in the reconstructions obtained using the Rotation (c) and Dual Scaling (d) predictors. They appear to be more resolved and separated compared to the reconstructions that used the Proximal (old) predictor (b). Additionally, the intensity levels within the two elliptical shapes more closely resemble those observed in the static reconstruction in Figure 4.

Figure 9 displays the PET reconstructions of the brain phantom. The initial frames are omitted for similar reasons. Without any predictor, the results are again worst for intervals where the phantom is rotating. Although the differences between the other reconstructions are less pronounced, a closer look reveals the subtle differences in intensities especially in the yellow region. When there is no motion, the reconstruction is best with the Dual Scaling predictor, and noisiest with the Zero Dual.

Figures 6 and 7 show the evolution of the SSIM and PSNR image quality metrics for the Shepp-Logan and brain phantoms, respectively. In both cases, the Proximal (old) predictor achieves lower SSIM and PSNR values compared to our proposed Greedy, Rotation, and Dual Scaling predictors. The image quality scores of brain phantom reconstructions with Strict Greedy predictor are better than those of the Proximal (old) predictor. Again, the reconstructions produced by the Dual Scaling predictor attained the highest SSIM and PSNR scores on intervals where rotational motion is stopped. Despite the true angles of rotation being zeroed out, the Dual Scaling predictor effectively adjusts for small errors in the available angles of rotation. Tables 2 and 3 additionally show average, SSIM and PSNR, as well as 95% confidence intervals (CI) for them. The Dual Scaling predictor has consistently the best performance according to these metrics. The Zero Dual, however, fares poorly on the SSIM.

CONCLUSIONS

The theory and the overall algorithm that we presented are simpler than the earlier approach of [33]. Moreover, the new predictors—justified by the theory—provide better numerical results. So far, we have, however, only treated relatively simple linear inverse problems with simple temporal characteristics. The next step will be: nonlinear inverse problems with realtime PDE solution based on the techniques of [18].

REFERENCES

[1] E. V. Belmega, P. Mertikopoulos, R. Negrel, and L. Sanguinetti, Online convex optimization and no-regret learning: Algorithms, guarantees and applications, 2018, arXiv:804.04529.

[2] M. A. Belzunce and A. J. Reader, Technical note: ultra high-resolution radiotracer-specific digital pet brain phantoms based on the BigBrain atlas, *Medical Physics* 47 (2020), 3356–3362, doi:10.1002/mp.14218.

[3] M. Benning, L. Gladden, D. Holland, C. B. Schönlieb, and T. Valkonen, Phase reconstruction from velocity-encoded MRI measurements – A survey of sparsity-promoting variational approaches, *Journal of Magnetic Resonance* 238 (2014), 26–43, doi:10.1016/j.jmr.2013.10.003.

[4] A. Bernstein, E. Dall'Anese, and A. Simonetto, Online primal-dual methods with measurement feedback for time-varying convex optimization, *IEEE Transactions on Signal Processing* 67 (2019), 1978–1991.

[5] A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans, Maximum-Likelihood Joint Image Reconstruction/Motion Estimation in Attenuation-Corrected Respiratory Gated PET/CT Using a Single Attenuation Map, *IEEE Transactions on Medical Imaging* 35 (2016), 217–228, doi:10.1109/tmi.2015.2464156.

[6] M. Burger, H. Dirks, L. Frerking, A. Hauptmann, T. Helin, and S. Siltanen, A variational reconstruction method for undersampled dynamic x-ray tomography based on physical motion models, *Inverse Problems* 33 (2017), 124008.

[7] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical Imaging and Vision* 20 (2004), 89–97, doi:10.1023/b:jmiv.0000011325.36760.1e.

[8] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision* 40 (2011), 120–145, doi:10.1007/s10851-010-0251-1.

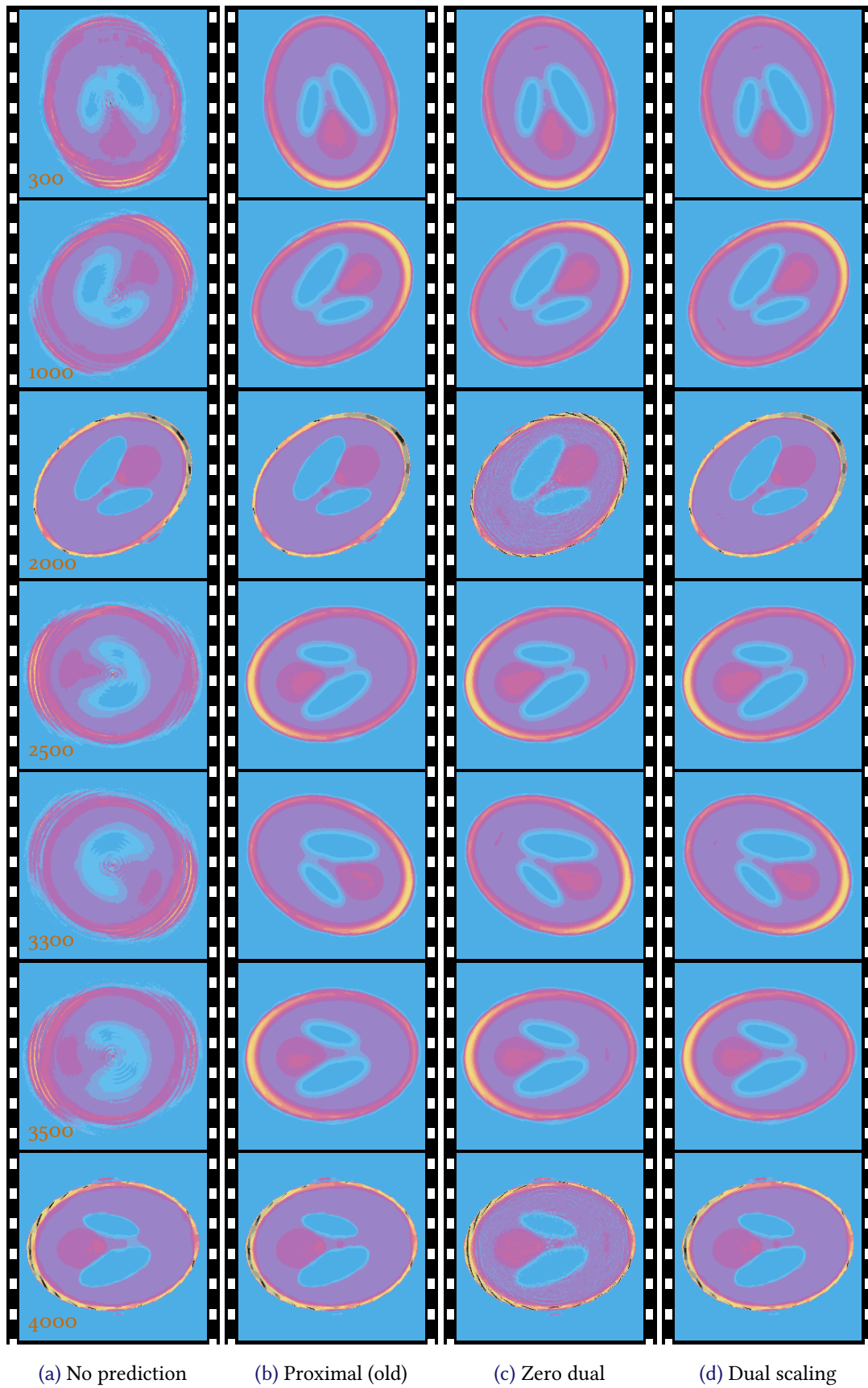(a) No prediction   (b) Proximal (old)   (c) Zero dual   (d) Dual scaling

Figure 8: Shepp–Logan reconstructions for several predictors. The colours represent values in $[0, 1]$ as ▮▮▮▮▮▮▮▮▮. The numbers on the left indicate the iteration.

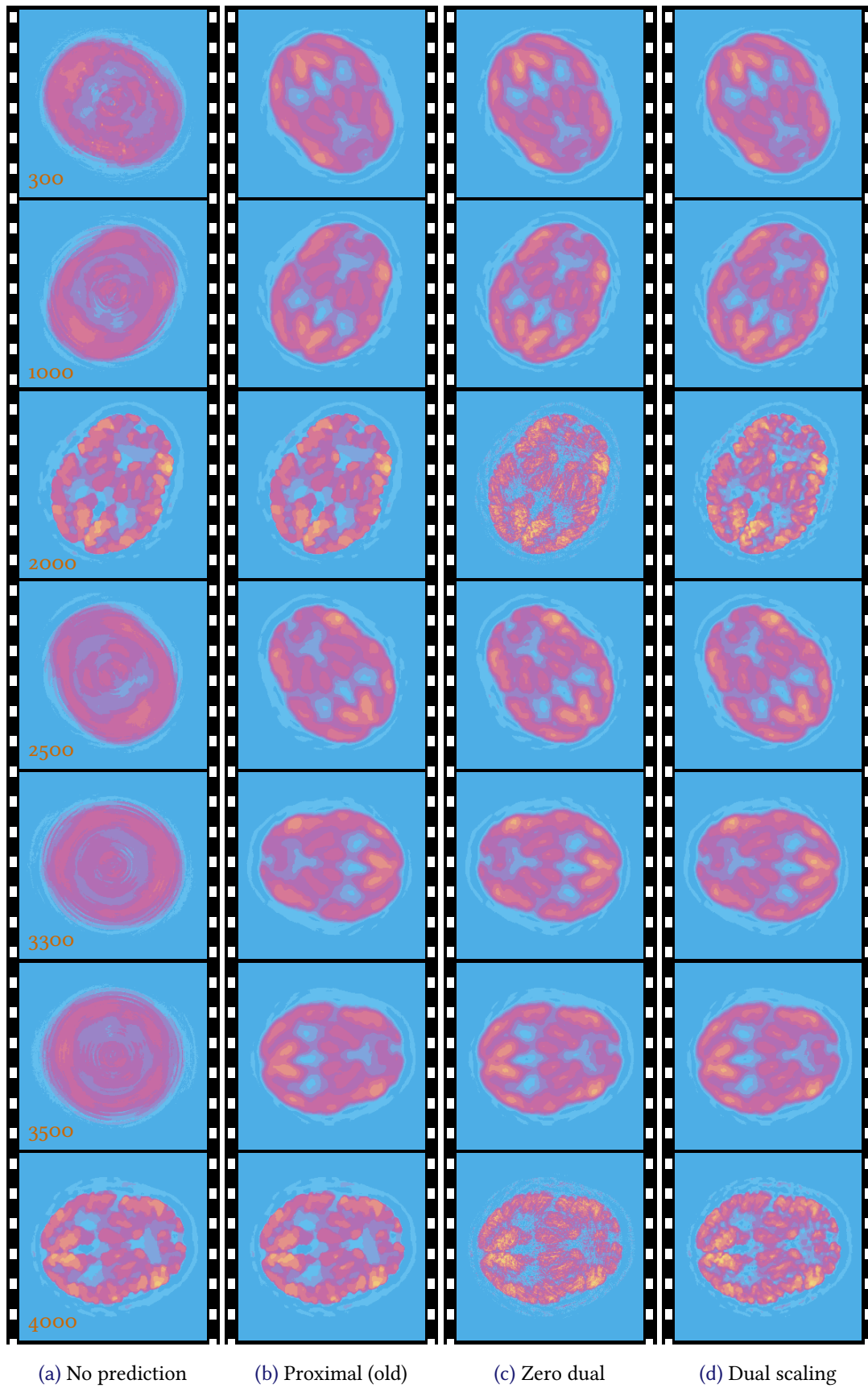(a) No prediction   (b) Proximal (old)   (c) Zero dual   (d) Dual scaling

Figure 9: Brain phantom reconstructions for several predictors. The colours represent values in $[0, 1]$ as ▮▮▮▮▮▮▮. The numbers on the left indicate the iteration.

[9]   T. J. Chang and S. Shahrampour, On online optimization: Dynamic regret analysis of strongly convex and smooth problems, in *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2021, 6966–6973.

[10] C. Clason and T. Valkonen, Introduction to Nonsmooth Analysis and Optimization, 2020, arXiv: 2001.00216. Work in progress.

[11] R. Franzen, Kodak lossless true color image suite, PhotoCD PCD0992. Lossless, true color images released by the Eastman Kodak Company, 1999, http://rok.us/graphics/kodak/.

[12] E. Hall and R. Willett, Dynamical models and tracking regret in online convex programming, in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester (eds.), volume 28 of Proceedings of Machine Learning Research, PMLR, Atlanta, Georgia, USA, 2013, 579–587, http://proceedings.mlr.press/v28/hall13.html.

[13] E. Hazan et al., Introduction to online convex optimization, *Foundations and Trends in Optimization* 2 (2016), 157–325.

[14] B. He and X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, *SIAM Journal on Imaging Sciences* 5 (2012), 119–149, doi:10.1137/100814494.

[15] D. J. Holland, D. M. Malioutov, A. Blake, A. J. Sederman, and L. F. Gladden, Reducing data acquisition times in phase-encoded velocity imaging u sing compressed sensing, *Journal of Magnetic Resonance* 203 (2010), 236–46.

[16] A. Hunt, Weighing without touching: applying electrical capacitance tomography to mass flowrate measurement in multiphase flows, *Measurement and Control* 47 (2014), 19–25, doi:10.1177/0020294013517445.

[17] Y. Iwao, G. Akamatsu, H. Tashima, M. Takahashi, and T. Yamaya, Brain PET motion correction using 3D face-shape model: the first clinical study, *Annals of Nuclear Medicine* 36 (2022), 904–912.

[18] B. Jensen and T. Valkonen, A nonsmooth primal-dual method with interwoven PDE constraint solver, *Computational Optimization and Appplications* (2024), doi:10.1007/s10589-024-00587-3, arXiv:2211.04807.

[19] S. Kar and J. M. Moura, Gossip and distributed Kalman filtering: Weak consensus under weak detectability, *IEEE Transactions on Signal Processing* 59 (2010), 1766–1784.

[20] A. Lipponen, A. Seppänen, and J. P. Kaipio, Nonstationary approximation error approach to imaging of three-dimensional pipe flow: experimental evaluation, *Measurement Science and Technology* 22 (2011), 104013, doi:10.1088/0957-0233/22/10/104013.

[21] F. Natterer, *The Mathematics of Computerized Tomography*, Society for Industrial and Applied Mathematics, 2001, doi:10.1137/1.9780898719284.

[22] M. Nonhoff and M. A. Müller, Online Gradient Descent for Linear Dynamical Systems, *IFAC-PapersOnLine* 53 (2020), 945–952, doi:https://doi.org/10.1016/j.ifacol.2020.12.1258. 21st IFAC World Congress.

[23] R. Olfati-Saber, J. A. Fax, and R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proceedings of the IEEE* 95 (2007), 215–233.

[24] F. Orabona, A Modern Introduction to Online Learning, 2020, arXiv:1912.13213.

[25] A. Simonetto, Dual prediction–correction methods for linearly constrained time-varying convex programs, *IEEE Transactions on Automatic Control* 64 (2018), 3355–3361.

[26] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, Time-varying convex optimization: Time-structured algorithms and applications, *Proceedings of the IEEE* 108 (2020), 2032–2048.

[27] A. Simonetto and E. Dall'Anese, Prediction-correction algorithms for time-varying constrained optimization, *IEEE Transactions on Signal Processing* 65 (2017), 5481–5494.

[28] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro, A class of prediction-correction methods for time-varying convex optimization, *IEEE Transactions on Signal Processing* 64 (2016), 4576–4591.

[29] Y. Tang, E. Dall'Anese, A. Bernstein, and S. Low, Running primal-dual gradient method for time-varying nonconvex problems, *SIAM Journal on Control And Optimization* 60 (2022), 1970–1990.

[30] M. Tico, Digital image stabilization, *Recent Advances in Signal Processing* (2009).

[31] T. Valkonen, Testing and non-linear preconditioning of the proximal point method, *Applied Mathematics and Optimization* 82 (2020), 591–636, doi:10.1007/s00245-018-9541-6.

[32] T. Valkonen, First-order primal-dual methods for nonsmooth nonconvex optimisation, in *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging*, K. Chen, C. B. Schönlieb, X. C. Tai, and L. Younes (eds.), Springer, Cham, 2021, doi:10.1007/978-3-030-03009-4_93-1, arXiv:1910.00115.

[33] T. Valkonen, Predictive online optimisation with applications to optical flow, *Journal of Mathematical Imaging and Vision* 63 (2021), 329–355, doi:10.1007/s10851-020-01000-4, arXiv:2002.03053.

[34] T. Valkonen, N. Dizon, and J. Jauhiainen, Predictive online optimisation codes for dynamic inverse imaging problems, Software on Zenodo, 2024, doi:10.5281/zenodo.12667014.

[35] L. Zhang, H. Liu, and X. Xiao, Regrets of proximal method of multipliers for online non-convex optimization with long term constraints, *Journal of Global Optimization* 85 (2023), 61–80.

[36] Y. Zhang, E. Dall'Anese, and M. Hong, Online proximal-ADMM for time-varying constrained convex optimization, *IEEE Transactions on Signal and Information Processing over Networks* 7 (2021), 144–155.

[37] Y. Zhang, R. J. Ravier, V. Tarokh, and M. M. Zavlanos, Distributed Online Convex Optimization with Improved Dynamic Regret, 2019, arXiv:1911.05127.

[38] J. Zhou, P. Hubel, M. Tico, A. N. Schulze, and R. Toft, Image registration methods for still image stabilization, 2016. US Patent 9,384,552.