Multiscale entropic regularization for MTS on general metric spaces

Farzam Ebrahimnejad* James R. Lee†

Paul G. Allen School of Computer Science & Engineering
University of Washington

Abstract

We present an $O((\log n)^2)$-competitive algorithm for metrical task systems (MTS) on any n-point metric space that is also 1-competitive for service costs. This matches the competitive ratio achieved by Bubeck, Cohen, Lee, and Lee (2019) and the refined competitive ratios obtained by Coester and Lee (2019). Those algorithms work by first randomly embedding the metric space into an ultrametric and then solving MTS there. In contrast, our algorithm is cast as regularized gradient descent where the regularizer is a multiscale metric entropy defined directly on the metric space. This answers an open question of Bubeck (Highlights of Algorithms, 2019).

Contents

*febrahim@cs.washington.edu

†jrl@cs.washington.edu

1 Introduction

Let (X, d) be a finite metric space with $|X| = n > 1$. The Metrical Task Systems (MTS) problem, introduced in [BLS92] is defined as follows. The input is a sequence $\langle c_t : X \to \mathbb{R}_+ \mid t = 1, 2, \ldots \rangle$ of nonnegative cost functions on the state space X. At every time t, an online algorithm maintains a state $\rho_t \in X$.

The corresponding cost is the sum of a *service cost* $c_t(\rho_t)$ and a *movement cost* $d(\rho_{t-1}, \rho_t)$. Formally, an *online algorithm* is a sequence of mappings $\boldsymbol{\rho} = \langle \rho_1, \rho_2, \ldots, \rangle$ where, for every $t \geqslant 1$, $\rho_t : (\mathbb{R}_+^X)^t \to X$ maps a sequence of cost functions $\langle c_1, \ldots, c_t \rangle$ to a state. The initial state $\rho_0 \in X$ is fixed. The *total cost of the algorithm* $\boldsymbol{\rho}$ *in servicing* $\boldsymbol{c} = \langle c_t : t \geqslant 1 \rangle$ is defined as the sum of the service and movement costs:

$$\begin{aligned}
\mathsf{serv}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \sum_{t \geqslant 1} c_t(\rho_t(c_1, \ldots, c_t)) \\
\mathsf{move}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \sum_{t \geqslant 1} d(\rho_{t-1}(c_1, \ldots, c_{t-1}), \rho_t(c_1, \ldots, c_t)) \\
\mathsf{cost}_{\boldsymbol{\rho}}(\boldsymbol{c}) &:= \mathsf{serv}_{\boldsymbol{\rho}}(\boldsymbol{c}) + \mathsf{move}_{\boldsymbol{\rho}}(\boldsymbol{c}).
\end{aligned}$$

The cost of the *offline optimum*, denoted $\mathsf{cost}^*(\boldsymbol{c})$, is the infimum of $\sum_{t \geqslant 1} [c_t(\rho_t) + d(\rho_{t-1}, \rho_t)]$ over *any* sequence $\langle \rho_t : t \geqslant 1 \rangle$ of states.

A *randomized online algorithm* $\boldsymbol{\rho}$ is said to be *α-competitive* if for every $\rho_0 \in X$, there is a constant $\beta > 0$ such that for all cost sequences $\boldsymbol{c}$:

$$\mathbb{E}\left[\mathsf{cost}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] \leqslant \alpha \cdot \mathsf{cost}^*(\boldsymbol{c}) + \beta\,.$$

Such an algorithm is said to be *α-competitive for service costs and α'-competitive for movement costs* if there is a constant $\beta > 0$ such that for all cost sequences $\boldsymbol{c}$:

$$\begin{aligned}
\mathbb{E}\left[\mathsf{serv}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] &\leqslant \alpha \cdot \mathsf{cost}^*(\boldsymbol{c}) + \beta \\
\mathbb{E}\left[\mathsf{move}_{\boldsymbol{\rho}}(\boldsymbol{c})\right] &\leqslant \alpha' \cdot \mathsf{cost}^*(\boldsymbol{c}) + \beta.
\end{aligned}$$

For the n-point uniform metric, a simple coupon-collector argument shows that the competitive ratio is $\Omega(\log n)$, and this is tight [BLS92]. A long-standing conjecture is that this $\Theta(\log n)$ competitive ratio holds for an arbitrary n-point metric space. The lower bound has almost been established [BBM06, BLMN05]; for any n-point metric space, the competitive ratio is $\Omega(\log n / \log\log n)$. Following a long sequence of works (see, e.g., [Sei99, BKRS00, BBBT97, Bar96, FM03, FRT04]), an upper bound of $O((\log n)^2)$ was shown in [BCLL21].

Competitive analysis via gradient descent. Let us consider an equivalent fractional perspective on MTS where the online algorithm maintains, at every point in time, a probability distribution $\mu_t \in \mathbb{R}_+^X$, and we interpret the costs similarly as a vector $c_t \in \mathbb{R}_+^X$. The cost of the algorithm is then given by

$$\sum_{t \geqslant 1} \left(\langle \mu_t, c_t \rangle + \mathbb{W}_X^1(\mu_{t-1}, \mu_t) \right),$$

where $\mathbb{W}_X^1$ is the L^1 transportation cost between two probability distributions on (X, d). This perspective is convenient, as now the state of the algorithm is given by a point in the probability simplex $\Delta_X \subseteq \mathbb{R}_+^X$.

This yields a natural first algorithm for solving MTS:

$$\mu_{t+1} := \mathrm{proj}_{\Delta_X}(\mu_t - \eta c_t), \tag{1.1}$$

where $\eta > 0$ is some parameter we can choose and proj_{Δ_X} denotes the Euclidean projection onto the convex body Δ_X. Moreover, it gives a natural way of relating the cost incurred by the algorithm to the cost incurred by *any other* state $\nu \in \Delta_X$: It is a basic exercise in convex geometry to show that

$$\|\mu_{t+1} - \nu\|^2 - \|\mu_t - \nu\|^2 \leqslant \eta\langle c_t, \nu - \mu_t\rangle. \tag{1.2}$$

In other words, if $\langle c_t, \mu_t\rangle > \langle c_t, \nu\rangle$, then μ_t approaches ν proportionally in the squared Euclidean distance.

Thus we cannot consistently incur more service cost than any fixed state. This does not provide a competitive algorithm because there is, in general, no convenient relationship between the Euclidean distance $\|\mu_t - \mu_{t+1}\|$ and the transportation distance $\mathbb{W}^1_X(\mu_t, \mu_{t+1})$.

But one can replace the Euclidean distance by any Bregman divergence $\mathbb{D}_\Phi$ associated to a strictly convex function Φ. Equivalently, we perform the projection (1.1) in the local inner product

$$\langle u, v\rangle_{\mu_t} := \langle \nabla^2\Phi(\mu_t)u, v\rangle.$$

Thus by choosing an appropriate geometry on Δ_X, one can hope to obtain a competitive algorithm. Such algorithms often go by the name *mirror descent* and the regularizer Φ is called the *mirror map* (we will often use the term *regularizer* interchangeably).

This framework is proposed in [ABBS10, BCN14] and applied to the k-server problem in [BCL+18], and to MTS in [BCLL21] and [CL19]. In all these papers, the algorithms apply only to ultrametrics (equivalently, to hierarchically separated tree metrics (HSTs)). In [BCLL21], mirror descent is used to analyze the algorithm on weighted stars, and these algorithms are glued together in an ad-hoc way to handle HSTs. In [CL19], stronger bounds (known as "refined guarantees") are obtained by finding an appropriate regularizer on arbitrary HSTs. In both cases, general finite metric spaces are then handled via random embeddings into HSTs.

In the present work, we apply this method directly to MTS on general metric spaces and match the best-known competitive ratio. Previously, it was unknown how to achieve any $\mathrm{poly}(\log n)$ competitive ratio for general metric spaces using mirror descent and achieving this was posed as an open problem by Bubeck[1].

We consider this an important step in advancing the underlying philosophy. Note that past approaches to MTS have involved a series of ad-hoc, complicated algorithms, along with clever potential function analyses. In contrast, in the mirror descent approach, once one specifies a convex body and a regularizer, both the algorithm and the method of analysis fall out naturally. Indeed, the most subtle part of competitive analysis lies in connecting the cost an online algorithm incurs to the cost of some offline optimum, and this is done entirely through the general Bregman divergence analog of (1.2), which becomes

$$\mathbb{D}_\Phi(\nu \,\|\, \mu_{t+1}) - \mathbb{D}_\Phi(\nu \,\|\, \mu_t) \leqslant \langle c_t, \nu - \mu_t\rangle.$$

[1]Posed in his talk at HALG 2019.

2 The multiscale noisy metric entropy

To obtain poly($\log n$)-competitive algorithms for MTS, previous approaches [BCLL21, CL19] employ a regularizer that can be cast as a multiscale entropy for probability distributions on an underlying tree metric. To handle general metric spaces, we will consider probability distributions on a lifted convex body that is specified by a directed ayclic graph whose sinks are the points of (X, d). See Figure 1 for a pictoral representation when the metric space is a path.

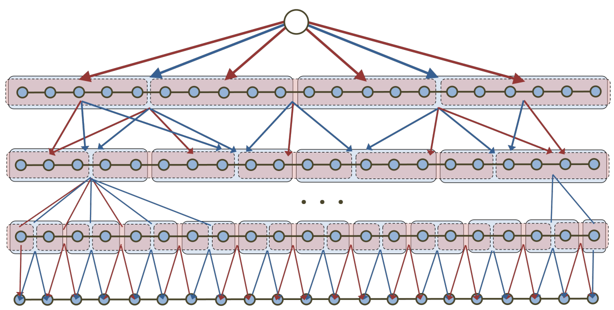

Figure 1: A hierarchical flow DAG over the path

The hierarchical flow DAG. Consider a finite set X and a directed ayclic weighted graph $\mathcal{D} = (V, A)$ with $X \subseteq V$ and such that

(i) $\mathcal{D}$ has a single source $\mathbb{r} \in V$, and

(ii) The set of sinks in $\mathcal{D}$ is X.

We say that $\mathcal{D}$ *is a DAG over* X. In what follows, we use the notation $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geqslant 0\}$ and $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$. For an arc $(u, v) \in A$, we will often use the shorthand uv.

A vector $F \in \mathbb{R}_+^A$ is called a *flow in* $\mathcal{D}$ if holds that

$$\sum_{v : uv \in A} F_{uv} = \sum_{v : vu \in A} F_{vu}, \quad \forall u \in V \setminus (X \cup \{\mathbb{r}\}). \tag{2.1}$$

For a flow F and $u \in V \setminus X$, define $F_u := \sum_{v : uv \in A} F_{uv}$. For a sink $x \in X$, we define $F_x := \sum_{u : ux \in A} F_{ux}$ as the flow into x. Say that F is a *unit flow in* $\mathcal{D}$ if $F_{\mathbb{r}} = 1$, and let $\mathcal{F}_{\mathcal{D}} \subseteq \mathbb{R}_+^A$ denote the convex set of all unit flows in $\mathcal{D}$.

A *(directed) path* γ *in* $\mathcal{D}$ is a sequence $\gamma = \langle u_1 u_2, u_2 u_3, \ldots, u_{m-1} u_m \rangle$ with $u_i u_{i+1} \in A$ for each $i \in \{1, \ldots, m-1\}$. We will occasionally also specify a path as a sequence of vertices. We use $\bar{\gamma}$ to denote the final vertex u_m of γ. Let $\mathcal{P}_{\mathcal{D}}$ denote the set of all paths in $\mathcal{D}$ from $\mathbb{r}$ to some sink.

The multiscale entropy. Let $\omega \in \mathbb{R}_{++}^A$ denote a vector of nonnegative arc lengths that are decreasing along paths, i.e., such that $\omega_{uv} > \omega_{vw}$ whenever $uv, vw \in A$. Let $\theta \in \mathbb{R}_{++}^A$ specify a probability distribution on the edges leaving every vertex, i.e.,

$$\sum_{v : uv \in A} \theta_{uv} = 1, \qquad \forall u \in V \setminus X. \tag{2.2}$$

Define the associated values

$$\eta_{uv} := 1 + \log(1/\theta_{uv}) \tag{2.3}$$
$$\delta_{uv} := \theta_{uv}/\eta_{uv}. \tag{2.4}$$

We refer to the triple $\hat{\mathcal{D}} := (\mathcal{D}, \omega, \theta)$ as a *marked DAG*. For a given normalization parameter $\kappa > 0$, such a marked DAG yields a multiscale entropy functional $\Phi_{\hat{\mathcal{D}}} : \mathcal{F}_{\mathcal{D}} \to \mathbb{R}_+$ defined by

$$\Phi_{\hat{\mathcal{D}}}(F) := \frac{1}{\kappa} \sum_{uv \in A} \frac{\omega_{uv}}{\eta_{uv}} \left(F_{uv} + \delta_{uv} F_u\right) \log\left(\frac{F_{uv}}{F_u} + \delta_{uv}\right).$$

One can consult [CL19] for a detailed discussion of multiscale entropies of this form on HSTs.

Two notions of depth. We define two notions of depth associated to $\hat{\mathcal{D}}$. The first is the *combinatorial depth* $\Delta_0(\mathcal{D})$ which is the maximum number of arcs in any path from $\mathbb{r}$ to some sink X. For $\gamma \in \mathcal{P}_{\mathcal{D}}$, let us define

$$\theta(\gamma) := \prod_{uv \in \gamma} \theta_{uv}, \tag{2.5}$$

and let the *information depth* be defined as

$$\Delta_I(\hat{\mathcal{D}}) := \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta(\gamma)).$$

Note that $\theta(\cdot)$ induces a probability distribution on $\mathcal{P}_{\mathcal{D}}$, and as clearly for $\gamma \in \mathcal{P}_{\mathcal{D}}$ it holds that $\theta(\gamma) \geqslant e^{-\Delta_I(\hat{\mathcal{D}})}$ we have

$$\log |\mathcal{P}_{\mathcal{D}}| \leqslant \Delta_I(\hat{\mathcal{D}}). \tag{2.6}$$

2.1 Mirror descent dynamics

Let us now fix a marked DAG $\hat{\mathcal{D}}$ and take $\Phi := \Phi_{\hat{\mathcal{D}}}$. We seek to define a continuous path $F : [0, \infty] \to \mathcal{F}_{\mathcal{D}}$ that represents the dynamics of projected vector flow in response to a continuous path $c(t) \in \mathbb{R}_+^X$ of costs arriving at the points of X.

A natural Euclidean flow would be specified heuristically by

$$F(t + dt) = \mathrm{proj}_{\mathcal{F}_{\mathcal{D}}} \left(F(t) - c(t)\, dt\right),$$

where for $v \in \mathbb{R}^A$, we define $\mathrm{proj}_{\mathcal{F}_{\mathcal{D}}}(v)$ as the unique point of $\mathcal{F}_{\mathcal{D}}$ with minimal Euclidean distance to v. In other words, we move a little in the direction $-c(t)$ and then project back to the feasible region $\mathcal{F}_{\mathcal{D}}$.

Instead, we will define our dynamics using the Bregman projection $\mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}}$ associated to our multiscale entropic regularizer, where

$$\mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}}(v) := \mathrm{argmin} \left\{\mathbb{D}_{\Phi}\left(v' \,\|\, v\right) : v' \in \mathcal{F}_{\mathcal{D}}\right\},$$

and

$$\mathbb{D}_{\Phi}\left(v' \,\|\, v\right) := \Phi(v') - \Phi(v) - \langle \nabla\Phi(v), v' - v\rangle$$

is the Bregman divergence associated to Φ.

One can show that if $c(t)$ is continuous, then there is a path $F : [0, \infty) \to \mathcal{F}_{\mathcal{D}}$ for which the following dynamics are well-defined (for almost every $t \in [0, \infty)$):

$$F(t + dt) = \mathrm{proj}^{\Phi}_{\mathcal{F}_{\mathcal{D}}} (F(t) - c(t)\, dt)$$

This path further satisfies (for almost all $t \in [0, \infty)$) the system of partial differential equations given by

$$\partial_t \left(\frac{F_{uv}(t)}{F_u(t)} \right) = \kappa \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv} \right) (\beta_u(t) - \hat{c}_{uv}(t)) \,, \quad uv \in A, \tag{2.7}$$

where $\hat{c}_{uv}(t) = \mathbb{1}_{\{F_{uv}(t)>0\}} c_v(t)$ if $v \in X$, and otherwise

$$\hat{c}_{uv}(t) = \mathbb{1}_{\{F_{uv}(t)>0\}} \sum_{w : vw \in A} \frac{F_{vw}(t)}{F_v(t)} \hat{c}_{vw}(t), \tag{2.8}$$

and $\beta_u(t)$ is the unique value that guarantees

$$\partial_t \sum_{v : uv \in A} \frac{F_{uv}(t)}{F_u(t)} = 0,$$

i.e.,

$$\beta_u(t) = \frac{\sum_{v : uv \in A} \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv} \right) \hat{c}_{uv}(t)}{\sum_{v : uv \in A} \frac{\eta_{uv}}{\omega_{uv}} \left(\frac{F_{uv}(t)}{F_u(t)} + \delta_{uv} \right)}.$$

Here we express the algorithm in continuous time for conceptual simplicity; its evolution is completely specified by the regularizer $\Phi_{\hat{\mathcal{D}}}$ and the costs $c(t)$. But the existence of a solution to (2.7) is derived from the limit of discrete-time algorithms in Section 4.

2.2 Metric compatibility

To analyze the algorithm specified by (2.7) on a metric space (X, d), we need additionally that $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ is compatible with the geometry of (X, d). Suppose that $\hat{\mathcal{D}}$ is a marked DAG over X. Say that *$\hat{\mathcal{D}}$ is τ-geometric* if it holds that for every pair of consecutive arcs $uv, vw \in A$, we have $\omega_{uv} \geqslant \tau \omega_{vw}$.

Let us define a metric on $\mathcal{P}_{\mathcal{D}}$ as follows: Suppose $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$ and let $u \in V$ be the first vertex at which they diverge, i.e., at which $uv_1 \in \gamma_1, uv_2 \in \gamma_2$ and $v_1 \neq v_2$. Define the distance

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) := \max(\omega_{uv_1}, \omega_{uv_2}).$$

One can check that this gives a metric on $\mathcal{P}_{\mathcal{D}}$ since the arc lengths are decreasing along source-sink paths. In fact, this defines an ultrametric on $\mathcal{P}_{\mathcal{D}}$.

Say that $\hat{\mathcal{D}}$ is *ε-expanding (with respect to (X, d))* if for every pair $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$, it holds that

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) \geqslant \varepsilon d(\bar{\gamma}_1, \bar{\gamma}_2),$$

where we recall that $\bar{\gamma}_1, \bar{\gamma}_2 \in X$ are the endpoints of γ_1 and γ_2, respectively.

We may extend $\mathrm{dist}_{\hat{\mathcal{D}}}$ to a distance on $\mathcal{F}_{\mathcal{D}}$ by defining $\mathbb{W}^1_{\hat{\mathcal{D}}}(F, F')$ as the L^1-transportation cost between $F, F' \in \mathcal{F}_{\mathcal{D}}$ with the underlying metric $\mathrm{dist}_{\hat{\mathcal{D}}}$, noting that F and F' can be viewed as probability distributions on $\mathcal{P}_{\mathcal{D}}$.

Say that $\hat{\mathcal{D}}$ *is L-Lipschitz (with respect to* (X, d)*)* if for every path $x_1, x_2, \ldots, x_m \in X$, there is a sequence of flows $F^{(1)}, F^{(2)}, \ldots, F^{(m)} \in \mathcal{F}_{\mathcal{D}}$ such that:

1. $F^{(i)}$ is a unit flow to x_i for every $i = 1, 2, \ldots, m$.
2. It holds that
$$\sum_{i=1}^{m-1} \mathbb{W}^1_{\hat{\mathcal{D}}}(F^{(i)}, F^{(i+1)}) \leqslant L \sum_{i=1}^{m-1} d(x_i, x_{i+1}).$$

Our main result follows from the next two theorems, which are proved in Section 4 and Section 3, respectively.

Theorem 2.1. *Suppose (X, d) is a metric space and $\hat{\mathcal{D}}$ is a τ-geometric marked DAG over X, for some $\tau \geqslant 4$. If $\hat{\mathcal{D}}$ is ε-expanding and L-Lipschitz with respect to (X, d), then for $\kappa = 6L$, the MTS algorithm specified by (2.7) is 1-competitive for service costs, and $O\left(\frac{L}{\varepsilon}\left(\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}})\right)\right)$-competitive for movement costs.*

Theorem 2.2. *For every n-point metric space (X, d), there is a 12-geometric marked DAG $\hat{\mathcal{D}}$ over X that is 1-expanding and $O(\log n)$-Lipschitz, and moreover satisfies*
$$\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}) \leqslant O(\log n).$$

3 Construction of a compatible DAG over (X, d)

In Section 3.1, we present the main construction of a marked DAG $\hat{\mathcal{D}}$ whose vertices are net points at every scale. Achieving the crucial property $\Delta_I(\hat{\mathcal{D}}) \leqslant O(\log n)$ requires choosing the net points and the arcs of $\mathcal{D}$ carefully. In Section 3.2, we argue that $\hat{\mathcal{D}}$ is ε-expanding and L-Lipschitz for $\varepsilon = 1$ and $L \leqslant O(\log n)$. It may not be that $\Delta_0(\mathcal{D}) \leqslant O(\log n)$, but in Section 3.3 we give a generic way of obtaining this property while leaving the other essential properties intact.

3.1 Hierarchical nets

Fix an n-point metric space (X, d) and assume, without loss of generality, that $\mathrm{diam}(X) = 1$. Define $\varepsilon := \min\{d(x, y) : x, y \in X\}$ and $K := 1 + \lceil \log_\tau(1/\varepsilon) \rceil$.

Construction of nets. Consider a parameter $\eta > 0$. We construct an η-net $N \subseteq X$ inductively as follows. Define $N_0 := \emptyset$ and for $j \geqslant 1$, inductively define the set
$$S_j := X \setminus B_X(N_{j-1}, \eta).$$
If $S_j = \emptyset$, then we take $N := N_{j-1}$. Otherwise, let $x_j \in S_j$ be a point that maximizes $|B_X(x, \eta/3)|$ among $x \in S_j$ and define $N_j := N_{j-1} \cup \{x_j\}$.

Lemma 3.1. *The set $N \subseteq X$ is an η-net with the property that for any set $W \subseteq X$, if*

$$x^* \in \operatorname{argmax}\{|B_X(y,\eta/3)| : y \in N \cap B_X(W, 1.5\eta)\},$$

then

$$|B_X(x^*,\eta/3)| \geqslant \max\{|B_X(w,\eta/3)| : w \in W\}$$

Proof. Suppose $x_j \in N$ is the element with j minimal such that $B_X(x_j, 1.5\eta) \cap W \neq \emptyset$. Then

$$\big(B_X(x_1,\eta) \cup \cdots \cup B_X(Q_{j-1},\eta)\big) \cap B_X(W,\eta/3) = \emptyset,$$

and hence by the greedy selection procedure,

$$\begin{aligned} &|B_X(x_j,\eta/3)| = |B_X(x^*,\eta/3)| \\ &|B_X(x_j,\eta/3)| \geqslant \max\{|B_X(w,\eta/3)| : w \in W\}, \end{aligned}$$

completing the proof. □

Denote $\tau := 12$. For each $k \in \{0,1,\ldots,K\}$, let U_k denote a τ^{-k}-net that satisfies Lemma 3.1 with $\eta = \tau^{-k}$. We now construct a DAG $\mathcal{D} = (V, A)$ with $V := \{(u,k) : u \in U_k, k \in \{0,1,\ldots,K\}\}$. For $k \in \{0,1,\ldots,K-1\}$, let A_k denote the collection of pairs (u,u') for every $u \in U_k$ and $u' \in U_{k+1}$ satisfying:

$$d(u,u') \leqslant 4\tau^{-k} \tag{3.1}$$

$$|B_X(u,\tau^{-k}/3)| \geqslant \max\left\{|B_X(w,\tau^{-k}/3)| : w \in B_X(u', 6\tau^{-(k+1)})\right\}. \tag{3.2}$$

We define $A := \bigcup_{k=0}^{K-1}\{((u,k),(u',k+1)) : (u,u') \in A_k\}$, and

$$\omega_{(u,k)(u',k+1)} := 10\tau^{-k}.$$

Since $U_K = X$, we can identify the sinks in $\mathcal{D}$ with the points of X. We take $\mathbb{r} := (u,0)$, where $U_0 = \{u\}$.

Observation 3.2. Suppose that $(u,k) \in U_k$ and $(x,K) \in V$ is reachable in $\mathcal{D}$ from (u,k). Then

$$d(u,x) \leqslant 4\tau^{-k} + 4\tau^{-(k+1)} + \cdots + 4\tau^{-(K-1)} < 5\tau^{-k}.$$

For a set $S \subseteq X$ and $k \in \{0,1,\ldots,K\}$, define

$$\varphi_k(S) := \operatorname{argmax}\left\{|B_X(y,\tau^{-k}/3)| : y \in B_X(S,2\tau^{-k}) \cap U_k\right\}. \tag{3.3}$$

We will require the following fact later.

Lemma 3.3. *Consider a set $S \subseteq X$ with $\operatorname{diam}_X(S) \leqslant 2\tau^{-k}$. If $u' \in S \cap U_{k+1}$, then $(\varphi_k(S), u') \in A_k$.*

Proof. Denote $u := \varphi_k(S)$. Since $u' \in S$ and $u \in B_X(S, 2\tau^{-k})$, it holds that $d(u,u') \leqslant 4\tau^{-k}$, and therefore (3.1) is satisfied. Now denote $W := B_X(u', 6\tau^{-(k+1)})$. Then $B_X(W, 1.5\tau^{-k}) \subseteq B_X(S, 2\tau^{-k})$, hence Lemma 3.1 implies that

$$|B_X(u,\tau^{-k}/3)| \geqslant \max\left\{|B_X(w,\tau^{-k}/3)| : w \in W\right\},$$

which shows that (3.2) is satisfied as well. □

For $k \in \{0, 1, \dots, K-1\}$ and $(u, u') \in A_k$, we define

$$\theta_{(u,k),(u',k+1)} := \frac{|B_X(u', \tau^{-(k+1)}/3)|}{\sum_{w:(u,w)\in A_k} |B_X(w, \tau^{-(k+1)}/3)|}. \tag{3.4}$$

Claim 3.4. It holds that

$$\sum_{w:(u,w)\in A_k} |B_X(w, \tau^{-(k+1)}/3)| \leqslant |B_X(u, 6\tau^{-k})|\,.$$

Proof. Since the elements of U_{k+1} form a $\tau^{-(k+1)}$-net, the balls $\left\{B_X(w, \tau^{-(k+1)}/3) : (u, w) \in A_k\right\}$ are pairwise disjoint. Furthermore, by Observation 3.2, every such ball is contained in

$$B_X(u, 5\tau^{-k} + \tau^{-(k+1)}/3) \subseteq B_X(u, 6\tau^{-k}).$$

□

Lemma 3.5. *It holds that $\Delta_I(\mathcal{D}) \leqslant 3\log n$, i.e., for every path $\gamma \in \mathcal{P}_{\mathcal{D}}$,*

$$\sum_{uv\in\gamma} \log(1/\theta_{u,v}) \leqslant 3\log n.$$

Proof. Consider a path $\gamma = \langle (u_0, 0), (u_1, 1), \dots, (u_K, K)\rangle$. From the definition (3.4) and Claim 3.4, it holds that

$$\sum_{k=0}^{K-1} \log\left(1/\theta_{(u_k,k)(u_{k+1},k+1)}\right) \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(u_k, 6\tau^{-k})|}{|B_X(u_{k+1}, \tau^{-(k+1)}/3)|}. \tag{3.5}$$

Let us denote $\ell := u_K$. By Observation 3.2, it holds that $d(\ell, u_k) \leqslant 6\tau^{-k}$ for $0 \leqslant k \leqslant K$. Therefore,

$$B_X(u_k, 6\tau^{-k}) \subseteq B_X(\ell, 12\tau^{-k}). \tag{3.6}$$

Furthermore since $(u_k, u_{k+1}) \in A_k$, by (3.2), we have

$$|B_X(u_k, \tau^{-k}/3)| \geqslant \max\left\{|B_X(w, \tau^{-k}/3)| : w \in B_X(u_{k+1}, 6\tau^{-(k+1)})\right\} \geqslant |B_X(\ell, \tau^{-k}/3)|, \tag{3.7}$$

since $d(\ell, u_{k+1}) \leqslant 6\tau^{-(k+1)}$.

By combining (3.5)–(3.7), we obtain

$$\sum_{k=0}^{K-1} \log\left(1/\theta_{(u_k,k)(u_{k+1},k+1)}\right) \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(\ell, 12\tau^{-k})|}{|B_X(\ell, \tau^{-(k+1)}/3)|} \leqslant \sum_{k=0}^{K-1} \log \frac{|B_X(\ell, \tau^{-(k-1)})|}{|B_X(\ell, \tau^{-(k+2)})|} \leqslant 3\log n,$$

where we used $\tau = 12$ in the penultimate inequality. □

The above result together with (2.6) yield the following.

Corollary 3.6. *It holds that $|\mathcal{P}_{\mathcal{D}}| \leqslant n^3$.*

3.2 Distortion analysis

Lemma 3.7. *It holds that $\hat{\mathcal{D}}$ is 1-expanding with respect to (X, d).*

Proof. Suppose that $\gamma_1, \gamma_2 \in \mathcal{P}_{\mathcal{D}}$ and let $u \in V$ be the first vertex for which $uv_1 \in \gamma_1$ and $uv_2 \in \gamma_2$ with $v_1 \neq v_2$. If $u = (x, k)$, then $\omega_{uv_1} = \omega_{uv_2} = 10\tau^{-k}$ and so $\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma_1, \gamma_2) = 10\tau^{-k}$. Moreover, by Observation 3.2 we have

$$d(\bar{\gamma}_1, \bar{\gamma}_2) \leqslant d(\bar{\gamma}_1, x) + d(\bar{\gamma}_2, x) \leqslant 10\tau^{-k},$$

completing the proof. □

For a partition P of X and $x \in X$, we let $P(x)$ denote the unique set in P containing X. We will require the following well-known random partitioning lemma.

Theorem 3.8 ([CKR01])**.** *For any finite metric space (X, d) and value $\Delta > 0$, there is a random partition P of X such that:*

1. $\mathrm{diam}_X(S) \leqslant \Delta$ *for every $S \in P$.*
2. *For all $x, y \in X$, it holds that*

$$\mathbb{P}\left[P(x) \neq P(y)\right] \leqslant 8\frac{d(x,y)}{\Delta} \log \frac{|B(x,\Delta)|}{|B(x,\Delta/8)|}.$$

For each $k \in \{0, 1, \ldots, K\}$, let P_k be a random partition of X satisfying the conclusion of Theorem 3.8 with $\Delta = \tau^{-k}$. Define a random map $\psi_k : X \to U_k$ as follows:

$$\psi_k(x) := \varphi_k(B_X(P_k(x), \tau^{-k}/2)),$$

where φ_k is the map defined in (3.3).

Lemma 3.9. *For every $x \in X$, it holds that $\langle(\psi_0(x), 0), (\psi_1(x), 1), \ldots, (\psi_K(x), K)\rangle$ is a path in $\mathcal{D}$.*

Proof. It suffices to show that for any $k \in \{0, 1, \ldots, K-1\}$, we have $(\psi_k(x), \psi_{k+1}(x)) \in A_k$. Define $u' = \psi_{k+1}(x)$ and $S := B_X(P_k(x), \tau^{-k}/2)$. Then $\mathrm{diam}_X(S) \leqslant 2\tau^{-k}$ and

$$d(x, u') = d(x, \psi_{k+1}(x)) \leqslant 2\tau^{-(k+1)} + \mathrm{diam}_X(B_X(P_{k+1}(x), \tau^{-(k+1)}/2)) \leqslant 4\tau^{-(k+1)} < \tau^{-k}/2,$$

where the last inequality follows from $\tau = 12$. Hence $u' \in S \cap U_{k+1}$. We can therefore apply Lemma 3.3 to conclude that $(\psi_k(x), u') = (\varphi_k(S), u') \in A_k$, completing the proof. □

For $x \in X$, define $\Psi(x) := \langle(\psi_0(x), 0), (\psi_1(x), 1), \ldots, (\psi_K(x), K)\rangle$. From the preceding lemma, we know that $\Psi : X \to \mathcal{P}_{\mathcal{D}}$.

Lemma 3.10. *For any $x, y \in X$, it holds that*

$$\mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x), \Psi(y))\right] \leqslant O(\log n)\, d(x, y).$$

Proof. From Theorem 3.8, we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x),\Psi(y))\right] &\leqslant \sum_{k=0}^{K} \mathbb{P}[P_k(x) \neq P_k(y)] \cdot 10\tau^{-k} \\
&\leqslant 80\, d(x,y) \sum_{k=0}^{K} \log \frac{|B(x,\tau^{-k})|}{|B(x,\tau^{-k}/8)|} \\
&\leqslant 80 \log(n)\, d(x,y),
\end{aligned}$$

where in the last line we used $\tau = 12 \geqslant 8$. □

Corollary 3.11. *It holds that $\hat{\mathcal{D}}$ is $O(\log n)$-Lipschitz with respect to (X,d).*

Proof. Consider any sequence $x_1,\dots,x_m$, and let us map it to the random sequence $\Psi(x_1),\dots,\Psi(x_m)$. Then from Lemma 3.10, we conclude

$$\sum_{j=1}^{m-1} \mathbb{E}\left[\mathrm{dist}_{\hat{\mathcal{D}}}(\Psi(x_j),\Psi(x_{j+1}))\right] \leqslant O(\log n) \sum_{j=1}^{m-1} d(x_j,x_{j+1}).$$

Hence there is a mapping $f : X \to \mathcal{P}_{\mathcal{D}}$ (that depends on the sequence $x_1,\dots,x_m$) such that $\sum_{j=1}^{m-1} d(f(x_j),f(x_{j+1})) \leqslant O(\log n) \sum_{j=1}^{m-1} d(x_j,x_{j+1})$, completing the proof. □

3.3 Compression

Let $\hat{\mathcal{D}} = (\mathcal{D},\omega,\theta)$ be the τ-geometric marked DAG constructed in Section 3.1. For a point $u \in V$, we let $\sigma(u)$ denote the number of paths in $\mathcal{D}$ that start at u and end in a point of X.

Observation 3.12. For $u \in V \setminus X$, it holds that

$$\sigma(u) = \sum_{v:uv\in A} \sigma(v). \tag{3.8}$$

Say an edge $uv \in A$ is *heavy* if $v \notin X$ and $\sigma(v) > \sigma(u)/2$; otherwise we say that uv *is light*. Moreover, we say a path $\gamma = \langle u_1u_2, u_2u_3, \dots, u_{m-1}u_m\rangle$ in $\mathcal{D}$ is *heavy-light* if all the edges $u_1u_2, u_2u_3, \dots, u_{m-2}u_{m-1}$ are heavy and $u_{m-1}u_m$ is light. The next lemma is straightforward and follows from (3.8).

Lemma 3.13. *For every $u \in V$, there is at most one heavy edge in $\mathcal{D}$ leaving u.*

Now we construct the marked DAG $\tilde{\mathcal{D}} = (\mathcal{D}',\omega',\theta')$ with $\mathcal{D}' = (V,A')$ as follows. We connect $u_i = (x_i,i) \in V$ to $u_j = (x_j,j) \in V$ for $1 \leqslant i < j \leqslant K$ in $\mathcal{D}'$ if there is a heavy-light path $\gamma = \langle u_iu_{i+1}, u_{i+1}u_{i+2}, \dots, u_{j-1}u_j\rangle$ from u_i to u_j in $\mathcal{D}$. Note that by Lemma 3.13, at most one such path can exist. We further set

$$\begin{aligned}
\omega'_{u_iu_j} &:= 10\tau^{-j+1}, \\
\theta'_{u_iu_j} &:= \prod_{k=i}^{j-1} \theta_{u_ku_{k+1}}.
\end{aligned}$$

Lemma 3.14. *For $uv \in A'$ with $v \notin X$ it holds that*

$$\sigma(v) \leqslant \sigma(u)/2.$$

Proof. Since $uv \in A'$, there must be a heavy-light path $\gamma = \langle w_1 w_2, \ldots, w_{m-1} w_m \rangle$ in $\mathcal{D}$ with $w_1 = u$ and $w_m = v$. Clearly the values of $\sigma(\cdot)$ are non-increasing along the (directed) paths in $\mathcal{D}$, hence we have

$$\sigma(u) \geqslant \sigma(w_2) \geqslant \cdots \geqslant \sigma(w_{m-1}).$$

Furthermore, as $w_{m-1}v$ is a light edge and $v \notin X$, it follows that

$$\sigma(v) \leqslant \sigma(w_{m-1})/2 \leqslant \sigma(u)/2,$$

as desired. □

Lemma 3.15. *It holds that $\Delta_0(\mathcal{D}') \leqslant O(\log |\mathcal{P}_{\mathcal{D}}|)$.*

Proof. We will argue that for every $\gamma \in \mathcal{P}_{\mathcal{D}'}$, one has $|\gamma| = O(\log |\mathcal{P}_{\mathcal{D}}|)$. Let $\gamma = \langle u_1 u_2, \ldots, u_{m-1} u_m \rangle$. Lemma 3.14 implies that for $1 \leqslant i \leqslant m-2$ we have $\sigma(u_{i+1}) \leqslant \sigma(u_i)/2$. Further note that we have $\sigma(u_1) = |\mathcal{P}_{\mathcal{D}}|$, and also clearly $\sigma(u_{m-1}) \geqslant 1$. Therefore,

$$m - 2 \leqslant \log_2(|\mathcal{P}_{\mathcal{D}}|),$$

completing the proof. □

We now define the map $f : \mathcal{P}_{\mathcal{D}} \to \mathcal{P}_{\mathcal{D}'}$ as follows. For a path $\gamma \in \mathcal{P}_{\mathcal{D}}$, let $f(\gamma)$ denote the path obtained by contracting all the heavy edges in γ. More precisely, for $\gamma = \langle u_1 u_2, u_2 u_3, \ldots, u_{m-1} u_m \rangle$, we define $f(\gamma)$ as follows. Denote $i_0 := 1$, and for $j = 1, 2, \ldots, m'$, let i_j denote the jth index for which $u_{i_j - 1} u_{i_j}$ is a light edge. We then denote

$$f(\gamma) := \langle u_{i_0} u_{i_1}, u_{i_1} u_{i_2}, \ldots, u_{i_{m'-1}} u_{i_{m'}} \rangle.$$

Lemma 3.16. *It holds that $\Delta_I(\tilde{D}) \leqslant \Delta_I(\hat{D})$.*

Proof. As all the edges in $\mathcal{D}'$ correspond to a path in $\mathcal{D}$, f is a surjective map. Furthermore, for $\gamma \in \mathcal{P}_{\mathcal{D}}$, one has $\theta(\gamma) = \theta'(f(\gamma))$, for $\theta(\cdot)$ defined as in (2.5) and $\theta'(\cdot)$ defined analogously, and thus we have

$$\Delta_I(\tilde{\mathcal{D}}) = \max_{\gamma' \in \mathcal{P}_{\mathcal{D}'}} \log(1/\theta'(\gamma')) = \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta'(f(\gamma))) \leqslant \max_{\gamma \in \mathcal{P}_{\mathcal{D}}} \log(1/\theta(\gamma)) = \Delta_I(\hat{D}),$$

completing the proof. □

Lemma 3.17. *For all $\gamma, \gamma' \in \mathcal{P}_{\mathcal{D}}$ it holds that*

$$\mathrm{dist}_{\hat{\mathcal{D}}}(\gamma, \gamma') = \mathrm{dist}_{\tilde{\mathcal{D}}}(f(\gamma), f(\gamma')).$$

Proof. Denote $\gamma = \langle u_1u_2, \ldots, u_{m-1}u_m \rangle$ and $\gamma' = \langle u_1'u_2', \ldots, u_{m-1}'u_m' \rangle$, and let $u_i = u_i'$ be the first vertex at which γ and γ' diverge so that we have

$$\mathrm{dist}_{\widehat{\mathcal{D}}}(P, P') = \max(\omega_{u_iu_{i+1}}, \omega_{u_iu_{i+1}'}) = 10\tau^{-i}.$$

By Lemma 3.13, at most one of u_iu_{i+1} and u_iu_{i+1}' can be heavy. Suppose that u_iu_{i+1} is light. Take $j := 1$ when $i = 1$, and otherwise let $j \leqslant i$ be the maximum index for which $u_{j-1}u_j$ is light. Further let $k \geqslant i$ be the minimum index for which $u_k'u_{k+1}'$ is a light edge. Note that k is well-defined because $u_{m-1}'u_m'$ is light. Now we have

$$\begin{aligned}\mathrm{dist}_{\tilde{\mathcal{D}}}(f(\gamma), f(\gamma')) &= \max(\omega'_{u_ju_{i+1}}, \omega'_{u_ju_{k+1}'})\\ &= \max(\omega_{u_iu_{i+1}}, \omega_{u_k'u_{k+1}'}) = \max(10\tau^{-i}, 10\tau^{-k}) = 10\tau^{-i},\end{aligned}$$

as desired. ◻

Lemma 3.18. *The $\tilde{D}$ is a marked DAG that is also τ-geometric.*

Proof. We first establish the τ-geometric property. Consider $u, v, w \in V$ with $uv, vw \in A'$. Denote $v = (x, i)$ for some $1 \leqslant i \leqslant K$. Then by construction, we have $\omega'_{uv} \geqslant 10\tau^{-i+1}$ and $\omega'_{vw} \leqslant 10\tau^{-i}$, completing the proof.

Next, we show that $\tilde{D}$ is a properly-constructed marked DAG. We need to establish that for $u \in V \setminus X$ it holds that

$$\sum_{v:uv\in A'} \theta'_{uv} = 1. \tag{3.9}$$

Let $v_0 := u$ and let $\gamma = \langle uv_1, v_1v_2, \ldots, v_{k-1}v_k \rangle$ be the maximal heavy path going out of u for some $k \geqslant 0$, meaning that all the edges v_iv_{i+1} are heavy for $0 \leqslant i \leqslant k-1$. Lemma 3.13 implies that the choice of γ is unique.

Now using (3.9), write

$$\begin{aligned}\sum_{v\,:\,uv\in A'} \theta'_{uv} &= \sum_{j=0}^{k-1} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-1} \theta_{v_\ell v_{\ell+1}} \cdot \left(\sum_{y\,:\,v_ky\in A} \theta_{v_ky} \right)\\ &= \sum_{j=0}^{k-1} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-1} \theta_{v_\ell v_{\ell+1}}\\ &= \sum_{j=0}^{k-2} \sum_{\substack{y\neq v_{j+1}:\\ v_jy\in A}} \theta_{v_jy} \cdot \prod_{\ell=0}^{j-1} \theta_{v_\ell v_{\ell+1}} + \prod_{\ell=0}^{k-2} \theta_{v_\ell v_{\ell+1}} \cdot \left(\sum_{y\,:\,v_{k-1}y\in A} \theta_{v_{k-1}y} \right)\\ &\;\;\vdots\\ &= \sum_{\substack{y\neq v_1:\\ v_0y\in A}} \theta_{v_0y} + \theta_{v_0v_1} \cdot \left(\sum_{y\,:\,v_1y\in A} \theta_{v_1y} \right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y\,:\,v_0 y\in A} \theta_{v_0 y} \\
&= 1,
\end{aligned}
$$

as desired. $\square$

We are now ready to prove Theorem 2.2.

Proof of Theorem 2.2. Let us show that $\tilde{\mathcal{D}}$ satisfies the requirements of the theorem. Lemma 3.18 shows that $\tilde{\mathcal{D}}$ is a 12-geometric marked DAG.

Now note that for $\gamma \in \mathcal{P}_{\mathcal{D}}$ we have $\overline{f(\gamma)} = \bar{\gamma}$, and thus Lemma 3.17 in conjunction with Lemma 3.7 and Corollary 3.11 implies that $\tilde{D}$ is 1-expanding and $O(\log n)$-Lipschitz. Moreover, Lemma 3.16 together with Lemma 3.5 bounds the information depth of $\tilde{D}$. Finally, a bound on the combinatorial depth follows from Lemma 3.15 and Corollary 3.6. $\square$

4 Algorithm and competitive analysis

4.1 Discrete-time algorithm

Let $\hat{\mathcal{D}} = (\mathcal{D}, w, \theta)$ be a marked DAG on X with $\mathcal{D} = (V, A)$. We now describe a generalization of the discrete-time dynamics of [CL19] on $\hat{\mathcal{D}}$ in response to a sequence of costs $\langle c_t : t \geqslant 1\rangle$, where $c_t \in \mathbb{R}_+^X$. Define

$$
Q_{\mathcal{D}} := \left\{ p \in \mathbb{R}_+^A \;\middle|\; \forall u \in V \setminus X : \sum_{v\,:\,uv\in A} p_{uv} = 1 \right\}.
$$

For $q \in Q_{\mathcal{D}}$ and $u \in V \setminus X$, we use $q^{(u)}$ to denote the restriction of q to the subspace spanned by subset of standard basis vectors $\{e_{uv} : uv \in A\}$, and we define the corresponding probability simplex $Q_{\mathcal{D}}^{(u)} := \{q^{(u)} : q \in Q_{\mathcal{D}}\}$. For convenience, we use $q_v^{(u)}$ for $q_{uv}^{(u)}$.

Let $\kappa > 0$ be a normalization parameter, and let the values η_{uv} and δ_{uv} be defined as in (2.3) and (2.4). For $u \in V \setminus X$ and $p \in Q_{\mathcal{D}}^{(u)}$, define

$$
\Phi^{(u)}(p) := \frac{1}{\kappa} \sum_{v\,:\,uv\in A} \frac{\omega_{uv}}{\eta_{uv}} (p_{uv} + \delta_{uv}) \log(p_{uv} + \delta_{uv}),
$$

and for $p' \in Q_{\mathcal{D}}^{(u)}$, denote

$$
\mathbb{D}^{(u)}(p \,\|\, p') := \mathbb{D}_{\Phi^{(u)}}(p \,\|\, p') = \frac{1}{\kappa} \sum_{v\,:\,uv\in A} \frac{\omega_{uv}}{\eta_{uv}} \left[(p_{uv} + \delta_{uv}) \log \frac{p_{uv} + \delta_{uv}}{p'_{uv} + \delta_{uv}} + p'_{uv} - p_{uv} \right].
$$

We now define an algorithm that takes a point $q \in Q_{\mathcal{D}}$ and a cost vector $c \in \mathbb{R}_+^X$ and outputs a point $p = \mathcal{A}(q, c) \in Q_{\mathcal{D}}$. Fix a topological ordering $\langle u_1, u_2, \ldots, u_N\rangle$ of $V \setminus X$ in $\mathcal{D}$. We define p inductively as follows. Denote $\hat{c}_x := c_x$ for $x \in X$, and for each $j = 1, 2, \ldots, N$:

$$
\hat{c}_v^{(u_j)} := \hat{c}_v \qquad \forall v : (u_j, v) \in A \tag{4.1}
$$

$$p^{(u_j)} := \operatorname{argmin}\left\{\mathbb{D}^{(u_j)}\left(p \,\|\, q^{(u_j)}\right) + \left\langle p, \hat{c}^{(u_j)}\right\rangle \;\middle|\; p \in Q_{\mathcal{D}}^{(u_j)}\right\} \tag{4.2}$$

$$\hat{c}_{u_j} := \sum_{v\,:\,u_j v \in A} p_v^{(u_j)}\, \hat{c}_v \tag{4.3}$$

We will use $\Lambda^{\mathcal{D}} : Q_{\mathcal{D}} \to \mathcal{F}_{\mathcal{D}}$ for the map which sends $q \in Q_{\mathcal{D}}$ to the (unique) $F = \Lambda^{\mathcal{D}}(q) \in \mathcal{F}_{\mathcal{D}}$ such that

$$F_{uv} = F_u q_{uv} \qquad \forall uv \in A.$$

Note that q contains more information than F; the map $\Lambda^{\mathcal{D}}$ fails to be invertible at $F \in \mathcal{F}_{\mathcal{D}}$ whenever there is some $u \in V \setminus X$ with $F_u = 0$. We will drop the superscript $\mathcal{D}$ from $\Lambda^{\mathcal{D}}$ whenever it is clear from context.

Now let p_0 be an arbitrary point in $Q_{\mathcal{D}}$. Given the cost sequence $\langle c_t : t \geqslant 1\rangle$, for $t = 1, 2, \ldots$ we define

$$p_t := \mathcal{A}(p_{t-1}, c_t), \tag{4.4}$$

and the associated MTS algorithm plays the distribution $\Lambda^{\mathcal{D}}(p_t)|_X$, i.e., at every $x \in X$ (recall that these are precisely the sinks in $\mathcal{D}$), the algorithm places probability mass equal to the flow in $\Lambda^{\mathcal{D}}(p_t)$ entering x.

For $c \in \mathbb{R}_+^X$ and $F \in \mathcal{F}_{\mathcal{D}}$ we define

$$\langle c, F\rangle_X := \sum_{v \in X} c_v F_v = \sum_{uv \in A\,:\,v \in X} c_v F_{uv}.$$

So the service cost of the algorithm until time $t \geqslant 1$ is given by

$$\sum_{s=1}^{t} \left\langle c, \Lambda^{\mathcal{D}}(p_s)\right\rangle_X,$$

and the movement cost is given by

$$\sum_{s=1}^{t} \mathbb{W}^1_{\hat{D}}\left(\Lambda^{\mathcal{D}}(p_{s-1}), \Lambda^{\mathcal{D}}(p_s)\right),$$

where we recall the L^1 transportation distance defined in Section 2.2.

4.2 Analysis via unfolding to an ultrametric

Let $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ be a τ-geometric marked DAG. As in Section 3.3, for a point $u \in V$, we define $\sigma(u)$ to denote the number of paths in $\mathcal{D}$ that start at u and end at X. Then if $\mathcal{D}$ is a tree and furthermore, for $uv \in A$, one defines

$$\theta_{uv} := \frac{\sigma(v)}{\sigma(u)}, \tag{4.5}$$

then the algorithm of the preceding section is exactly the same as the one for HSTs introduced in [CL19], as $\sigma(u)$ is precisely the number of leaves in the subtree rooted at u. The next result is a restatement of [CL19, Thm. 2.7].

Theorem 4.1 ([CL19])**.** *Let $\hat{\mathcal{D}} = (\mathcal{D}, \omega, \theta)$ be a τ-geometric marked DAG over X, for some $\tau \geqslant 4$, and such that $\mathcal{D}$ is a tree. If θ is defined as in (4.5), and $\hat{\mathcal{D}}$ is 1-expanding and L-Lipschitz, then there is some value $\kappa \asymp L$ and a number $\varepsilon = \varepsilon(\hat{\mathcal{D}}) > 0$ so that for any sequence of cost vectors $\langle c_t : t \geqslant 1\rangle$ satisfying $\|c_t\|_\infty \leqslant \varepsilon$, the MTS algorithm specified in Section 4.1 is 1-competitive for service costs and $O\big(L\left(\Delta_0(\mathcal{D}) + \log |X|\right)\big)$-competitive for movement costs.*

Note that the condition on the ℓ_∞ norm of the cost vectors in the above theorem is not restrictive, since as noted in [CL19], we can always split arbitrary cost vectors into smaller pieces with each satisfying the desired ℓ_∞ bound.

Our goal now is to show that if θ is defined as in (4.5), then similar guarantees as in Theorem 4.1 hold for the algorithm on $\hat{\mathcal{D}}$, even when $\mathcal{D}$ is not a tree.

Theorem 4.2. *Let $\hat{\mathcal{D}}$ be a τ-geometric marked DAG over X, for some $\tau \geqslant 4$, and such that θ is given by (4.5). If $\hat{\mathcal{D}}$ is 1-expanding and L-Lipschitz, then there is some value $\kappa \asymp L$ and a number $\varepsilon = \varepsilon(\hat{\mathcal{D}}) > 0$ so that for any sequence of cost vectors $\langle c_t : t \geqslant 1\rangle$ satisfying $\|c_t\|_\infty \leqslant \varepsilon$, the MTS algorithm specified in Section 4.1 is 1-competitive for service costs and $O\big(L\left(\Delta_0(\mathcal{D}) + \log |X|\right)\big)$-competitive for movement costs.*

Note that from (2.6) it follows that

$$\Delta_0(\mathcal{D}) + \log |\mathcal{P}_{\mathcal{D}}| \leqslant \Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}),$$

and hence the above theorem together with Theorem 2.2 already gives a competitive algorithm with our desired bounds, though only for the specific choice of θ given by (4.5). In Section 4.3, we address the case of general θ.

We prove Theorem 4.2 via a simple reduction to Theorem 4.1. Consider a τ-geometric marked DAG $\hat{\mathcal{D}} = (\mathcal{D}, w, \theta)$ on X with $\mathcal{D} = (V, A)$. Note that $d_{\hat{\mathcal{D}}}$ defines an ultrametric on $\mathcal{P}_{\mathcal{D}}$. We show that the dynamics on $\hat{\mathcal{D}}$ are "equivalent" to the dynamics on the HST corresponding to the ultrametric $(\mathcal{P}_{\mathcal{D}}, d_{\hat{\mathcal{D}}})$. More precisely, let us construct the τ-geometric marked tree $\tilde{\mathcal{D}} = (\mathcal{D}', w', \theta')$ with $\mathcal{D}' = (V', A')$ as follows. We define V' as the set of (directed) paths is $\mathcal{D}$ originating from the root. Furthermore, we connect $\gamma \in V'$ to $\gamma' \in V'$ whenever γ' is formed by adding the edge $\bar{\gamma}'$ to γ, and set

$$\omega_{\gamma\gamma'} := \omega_{\bar{\gamma}'}, \qquad \theta_{\gamma\gamma'} := \theta_{\bar{\gamma}'}.$$

One can verify that $\tilde{\mathcal{D}}$ is a τ-geometric marked tree over $\mathcal{P}_{\mathcal{D}}$. Moreover, since $\mathcal{D}'$ is a tree, there is a natural identification between the elements of $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{D}'}$ so that for $\gamma, \gamma' \in \mathcal{P}_{\mathcal{D}}$ it holds that

$$d_{\hat{\mathcal{D}}}(\gamma, \gamma') = d_{\tilde{\mathcal{D}}}(\gamma, \gamma'). \tag{4.6}$$

Now for $p \in \mathcal{Q}_{\mathcal{D}}$, define $\tilde{p} \in Q_{\mathcal{D}'}$ to be the natural extension of p in $\mathcal{D}'$ so that for $\gamma\gamma' \in A'$ one has $\tilde{p}_{\gamma\gamma'} = p_{\bar{\gamma}'}$. Furthermore, for a cost sequence $c \in \mathbb{R}_+^X$ define its extension $\tilde{c} \in \mathbb{R}^{\mathcal{P}_{\mathcal{D}}}$ as the vector with $\tilde{c}_\gamma = c_{\bar{\gamma}}$ for $\gamma \in \mathcal{P}_{\mathcal{D}}$. Finally, let $\mathcal{A}$ denote the single-step discrete dynamics on $\hat{\mathcal{D}}$ as defined in Section 4.1, and similarly let $\mathcal{A}'$ denote the discrete dynamics on $\tilde{\mathcal{D}}$. Then the following lemma is straightforward.

Lemma 4.3. *Let $p \in \mathcal{Q}_{\mathcal{D}}$, $c \in \mathbb{R}_+^X$. Then it holds that*

$$\langle \Lambda^{(\mathcal{D})}(p), c\rangle_X = \langle \Lambda^{(\mathcal{D}')}(\tilde{p}), \tilde{c}\rangle_{\mathcal{P}_{\mathcal{D}}}. \tag{4.7}$$

Furthermore, for $q = \mathcal{A}(p, c)$ we have

$$\mathcal{A}'(\tilde{p}, \tilde{c}) = \tilde{q}. \tag{4.8}$$

We are now ready to prove the main result of this section.

Proof of Theorem 4.2. Let $p_0 \in \mathcal{Q}_{\mathcal{D}}$ and $q_0 \in Q_{\mathcal{D}'}$ with $q_0 = \tilde{p_0}$. Given the cost sequence $\langle c_t : t \geqslant 1\rangle$, for $t \geqslant 1$ let

$$p_t = \mathcal{A}(p_{t-1}, c_t)$$

and

$$q_t = \mathcal{A}'(q_{t-1}, \tilde{c}_t).$$

Then by repeatedly applying (4.8) we get that for $t \geqslant 1$ we have $q_t = \tilde{p_t}$. Therefore from (4.7) and (4.6) it follows that the service and movement costs of the dynamics on $\hat{\mathcal{D}}$ and $\tilde{\mathcal{D}}$ are equal. Hence the competitiveness guarantees for the dynamics on $\hat{\mathcal{D}}$ follow from an application of Theorem 4.1 to the dynamics on $\tilde{\mathcal{D}}$, completing the proof. □

4.3 Analysis of the general case

We now prove Theorem 2.1 via a relatively straightforward generalization of the analysis in [CL19]. Let $\hat{\mathcal{D}} = (\mathcal{D}, w, \theta)$ be a τ-geometric marked DAG with $\tau \geqslant 4$, and consider the mirror descent dynamics on $\hat{\mathcal{D}}$ described in Section 4.1.

For a unit flow $F \in \mathcal{F}_{\mathcal{D}}$ and $q \in \mathcal{Q}_{\mathcal{D}}$, define the global divergence function

$$\mathbb{D}(F \,\|\, q) := \frac{1}{\kappa} \sum_{uv \in A} \frac{\omega_{uv}}{\eta_{uv}} \left[(F_{uv} + F_u \delta_{uv}) \log \frac{\frac{F_{uv}}{F_u} + \delta_{uv}}{q_{uv} + \delta_{uv}} + F_u q_{uv} - F_{uv} \right],$$

with the convention that $0 \log\left(\frac{0}{0} + \delta_v\right) = \lim_{\varepsilon \to 0} \varepsilon \log\left(\frac{0}{\varepsilon} + \delta_v\right) = 0$. We further define the norm $\ell_1(\omega)$ as

$$\|F\|_{\ell_1(\omega)} = \sum_{uv \in A} \omega_{uv} |F_{uv}|.$$

Observation 4.4. For $F, F' \in \mathcal{F}_{\mathcal{D}}$ it holds that

$$\frac{1}{2} \|F - F'\|_{\ell_1(\omega)} \leqslant \mathbb{W}^1_{\hat{\mathcal{D}}}(F, F') \leqslant \|F - F'\|_{\ell_1(\omega)}.$$

The next lemma lets us bound the amount of change of the global divergence when the offline algorithm makes a movement.

Lemma 4.5 ([CL19, Lemma 2.2])**.** *For flows $F, F' \in \mathcal{F}_{\mathcal{D}}$ and $q \in \mathcal{Q}_{\mathcal{D}}$ we have*

$$|\mathbb{D}(F \,\|\, q) - \mathbb{D}(F' \,\|\, q)| \leqslant \frac{1}{\kappa}(2 + \frac{4}{\tau}) \|F - F'\|_{\ell_1(\omega)}$$

Suppose $q \in \mathcal{Q}_{\mathcal{D}}$, $p = \mathcal{A}(q, c)$, and further let $Q = \mu(q)$, $P = \mu(p)$. The KKT conditions for (4.2) give: For every $uv \in A$,

$$\frac{1}{\kappa} \frac{\omega_{uv}}{\eta_{uv}} \log\left(\frac{p_{uv} + \delta_{uv}}{q_{uv} + \delta_{uv}}\right) = \beta_u - \hat{c}_v + \alpha_{uv}\,, \tag{4.9}$$

where α_{uv} is the Lagrange multipliers corresponding to the nonnegativity constraints in (4.2), $\beta_u \geqslant 0$ is the multiplier corresponding to the constraint $\sum_{v: uv \in A} q_{uv} \geqslant 1$, and $\hat{c}$ is defined as in (4.3).

Note that as in [CL19] the nonnegativity multipliers are unique and thus well-defined here. The complementary slackness conditions give us

$$\alpha_{uv} > 0 \implies p_{uv} = 0. \tag{4.10}$$

We use $\alpha^{(u)}$ to denote the restriction of α to the subspace spanned by $\{e_{uv} : uv \in A\}$.

The following two lemmas, which allow us to bound the service cost and the movement cost of the algorithm, respectively, are the main ingredients in the proof of Theorem 2.2.

Lemma 4.6. *It holds that*

$$\mathbb{D}(F \,\|\, p) - \mathbb{D}(F \,\|\, q) \leqslant \langle c, F - P \rangle_X .$$

Define $\omega_{\min} := \min_{uv \in A}\{\omega_{uv}\}$ and

$$\varepsilon_{\mathcal{D}} := \frac{\omega_{\min}}{2(2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}))} \frac{\tau - 3}{\tau\kappa}.$$

Furthermore, for $F \in \mathcal{P}_{\mathcal{D}}$ and $r \in \mathcal{Q}_{\mathcal{D}}$ define

$$\psi(F) := \sum_{uv \in A} \omega_{uv} F_{uv}$$

and

$$\begin{aligned} \Psi_u(r) &:= -\Lambda^{\mathcal{D}}(r)_u \, \mathbb{D}^{(u)}\left(\theta^{(u)} \,\|\, r^{(u)}\right) \\ \Psi(r) &:= \sum_{u \in V \setminus X} \Psi_u(r). \end{aligned}$$

Lemma 4.7. *For any $Z \in \mathcal{F}_{\mathcal{D}}$:*

$$\kappa^{-1} \|Q - P\|_{\ell_1(\omega)} \leqslant [\psi(Y) - \psi(X)] + \frac{2\tau}{\tau - 3}\left([\Psi(q) - \Psi(p)] + (2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}))\langle c, Q\rangle_X\right). \tag{4.11}$$

Moreover, if $\|c\|_\infty \leqslant \varepsilon_{\mathcal{D}}$, then

$$\kappa^{-1} \|Q - P\|_{\ell_1(\omega)} \leqslant [\psi(Y) - \psi(X)] + \frac{4\tau}{\tau - 3}\left([\Psi(q) - \Psi(p)] + (2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}}))\langle c, P\rangle_X\right). \tag{4.12}$$

We prove Lemma 4.6 and Lemma 4.7 in Section 4.4 and Section 4.5, respectively. Now given these results, let us prove Theorem 2.2.

Proof of Theorem 2.2. Consider a sequence $\langle c_t : t \geqslant 1\rangle$ of cost functions. By splitting the costs into smaller pieces, we may assume that $\|c_t\|_\infty \leqslant \varepsilon_{\mathcal{D}}$ for all $t \geqslant 1$.

Let $t_1 \geqslant 1$, and let $r_0^*, r_1^*, \ldots, r_{t_1}^* \in X$ denote the path taken by an (optimal) offline algorithm in response to the cost sequence $\langle c_t : t \geqslant 1\rangle$. The L-Lipschitzness property of $\hat{\mathcal{D}}$ implies that there exists a sequence $R_0^*, R_1^*, \ldots R_{t_1}^* \in \mathcal{F}_{\mathcal{D}}$ such that R_i^* is a unit flow to r_i, and furthermore

$$\sum_{i=1}^{t_1} \mathbb{W}^1_{\hat{\mathcal{D}}}(R_{i-1}^*, R_i^*) \leqslant L \sum_{i=1}^{t_1} d(r_{i-1}^*, r_i^*). \tag{4.13}$$

Let $q_0, \ldots, q_{t_1} \in \mathcal{Q}_{\mathcal{D}}$ denote the trajectory of the discrete mirror descent dynamics with $\kappa = 6L$ on $\hat{\mathcal{D}}$ in response to the cost sequence $\langle c_t : t \geqslant 1\rangle$. Further let $\{Q_t = \mu(q_t)\}$, and suppose $R_0^* = Q_0^*$. Then using $\mathbb{D}(R_0^* \,\|\, q_0) = 0$ along with Lemma 4.6 and Lemma 4.5 yields, for any time $t_1 \geqslant 1$,

$$\begin{aligned}
\sum_{t=1}^{t_1} \langle c_t, Q_t\rangle_X &\leqslant \sum_{t=1}^{t_1} \langle c_t, Z_t^*\rangle_X - \mathbb{D}(R_{t_1}^* \,\|\, q_{t_1}) + \frac{3}{\kappa} \sum_{t=1}^{t_1} \|R_t^* - R_{t-1}^*\|_{\ell_1(\omega)} \\
&\leqslant \sum_{t=1}^{t_1} \langle c_t, R_t^*\rangle_X + \frac{3}{\kappa} \sum_{t=1}^{t_1} \|R_t^* - R_{t-1}^*\|_{\ell_1(\omega)} \\
&\leqslant \sum_{t=1}^{t_1} \langle c_t, R_t^*\rangle_X + \frac{6L}{\kappa} \sum_{t=1}^{t_1} d(r_{t-1}^*, r_t^*),
\end{aligned}$$

where in the second line we have used $\mathbb{D}(R \,\|\, q) \geqslant 0$ for all $R \in \mathcal{F}_{\mathcal{D}}$ and $q \in \mathcal{Q}_{\mathcal{D}}$, and the last line follows from Observation 4.4 and (4.13). This confirms that the mirror descent dynamics is 1-competitive for the service costs. Now we can write

$$\begin{aligned}
\frac{\varepsilon}{\kappa} \sum_{t=1}^{t_1} \mathbb{W}_X^1(Q_{t-1}, Q_t) &\leqslant \frac{1}{\kappa} \sum_{t=1}^{t_1} \mathbb{W}_{\hat{\mathcal{D}}}^1(Q_{t-1}, Q_t) && (\hat{\mathcal{D}} \text{ is } \varepsilon\text{-expanding}) \\
&\leqslant \frac{1}{\varepsilon} \sum_{t=1}^{t_1} \|Q_t - Q_{t-1}\|_{\ell_1(\omega)} && (\text{Observation 4.4}) \\
&\leqslant [\psi(Q_{t_1}) - \psi(Q_0)] + \frac{4\tau}{\tau - 3}\left([\Psi(q_0) - \Psi(q_{t_1})] + \left(2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}})\right) \sum_{t=1}^{t_1} \langle c_t, Q_t\rangle_X\right),
\end{aligned}$$

where in the last line we used (4.12). This implies that the mirror descent dynamics is $(96L/\varepsilon) \cdot \left(2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}})\right)$-competitive in the movement cost, completing the proof. □

4.4 Bounding the service cost

In this section we prove Lemma 4.6. Let $F \in \mathcal{F}_{\mathcal{D}}$, and for $u \in V \setminus X$ with $F_u > 0$, define $F^{(u)} \in Q_{\mathcal{D}}^{(u)}$ by

$$F_v^{(u)} := \frac{F_{uv}}{F_u}.$$

The next lemma is a consequence of [CL19, Lemma 2.1].

Lemma 4.8. *For $u \in V \setminus X$ we have*

$$\mathbb{D}^{(u)}\left(F^{(u)} \,\|\, p^{(u)}\right) - \mathbb{D}^{(u)}\left(F^{(u)} \,\|\, q^{(u)}\right) \leqslant \left\langle \hat{c}^{(u)} - \alpha^{(u)}, F^{(u)} - p^{(u)} \right\rangle. \tag{4.14}$$

Proof of Lemma 4.6. Multiplying both sides of (4.14) by F_u and summing over all $u \in V \setminus X$ yields

$$\mathbb{D}(F \,\|\, p) - \mathbb{D}(F \,\|\, q) \leqslant \sum_{u \in V \setminus X} F_u \left\langle \hat{c}^{(u)} - \alpha^{(u)}, F^{(u)} - p^{(u)} \right\rangle$$

$$
\begin{aligned}
&= \sum_{uv\in A} F_{uv}(\hat{c}_v^{(u)} - \alpha_v^{(u)}) - \sum_{uv\in A} F_u p_{uv}(\hat{c}_v^{(u)} - \alpha_v^{(u)}) \\
&\leqslant \sum_{uv\in A} F_{uv}\hat{c}_v^{(u)} - \sum_{uv\in A} F_u p_{uv}(\hat{c}_v^{(u)} - \alpha_v^{(u)}).' \qquad (\alpha_v^{(u)} \geqslant 0)
\end{aligned}
$$

Note that from (4.10) the latter expression is

$$
\sum_{u\notin X} F_u \sum_{v:uv\in A} \hat{c}_v^{(u)} p_v = \sum_{u\notin X} F_u \hat{c}_u .
$$

Noting that $\hat{c}_{\mathbb{r}} = \sum_{u\in X} \mu(p)_u c_u$, this gives

$$
\mathbb{D}(F \parallel p) - \mathbb{D}(F \parallel q) \leqslant \sum_{u\neq \mathbb{r}} \hat{c}_u F_u - \sum_{u\in V\setminus X} F_u \hat{c}_u \leqslant \langle c, F - P\rangle_X . \qquad \square
$$

4.5 Bounding the the movement cost

In this section we prove Lemma 4.7. The next lemma shows that when the algorithm moves from Q to P it suffices for us to bound the positive movement movement cost $\|(P-Q)_+\|_{\ell_1(\omega)}$.

Lemma 4.9 ([CL19, Lemma 2.4])**.** *For $F, F' \in \mathcal{F}_{\mathfrak{D}}$ it holds that*

$$
\|F - F'\|_{\ell_1(\omega)} = 2\,\|(F-F')_+\|_{\ell_1(\omega)} + [\psi(F') - \psi(F)].
$$

Lemma 4.10 ([CL19, Lemma 2.9])**.** *It holds that $\alpha_{uv} \leqslant \hat{c}_v$ for all $uv \in A$.*

Define $\rho_{uv} := \log\left(\frac{p_{uv}+\delta_{uv}}{q_{uv}+\delta_{uv}}\right)$ so that

$$
q_{uv} - p_{uv} = (q_{uv} + \delta_{uv})(1 - e^{\rho_{uv}}). \tag{4.15}
$$

Recall that for $uv \in A$, we have $Q_{uv} = q_{uv}Q_u$ and $P_{uv} = p_{uv}P_u$, thus

$$
Q_{uv} - P_{uv} = Q_u(q_{uv} - p_{uv}) + p_{uv}(Q_u - P_u) = (Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}}) + p_{uv}(Q_u - P_u).
$$

In particular,

$$
\begin{aligned}
\omega_{uv}\,(Q_{uv} - P_{uv})_+ &\leqslant \omega_{uv}(Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}})_+ + \omega_{uv}p_{uv}\,(Q_u - P_u)_+ \\
&\leqslant \omega_{uv}(Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}})_+ + \sum_{w:wu\in A} \omega_{uv}p_{uv}\,(Q_{wu} - P_{wu})_+ \\
&\leqslant \omega_{uv}(Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}})_+ + \sum_{w:wu\in A} \frac{\omega_{wu}}{\tau} p_{uv}\,(Q_{wu} - P_{wu})_+ .
\end{aligned}
$$

Using $\sum_{v:uv\in A} p_{uv} = 1$ and summing over all edges yields

$$
\sum_{uv\in A} \omega_{uv}\,(Q_{uv} - P_{uv})_+ \leqslant \sum_{uv\in A} \omega_{uv}(Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}})_+ + \frac{1}{\tau}\sum_{uv\in A} \omega_{uv}\,(Q_{uv} - P_{uv})_+ \,,
$$

hence

$$
\sum_{uv\in A} \omega_{uv}\,(Q_{uv} - P_{uv})_+ \leqslant \frac{\tau}{\tau - 1}\sum_{uv\in A} \omega_{uv}(Q_{uv} + \delta_{uv}Q_u)(1 - e^{\rho_{uv}})_+
$$

$$\leqslant \frac{\tau}{\tau-1}\sum_{uv\in A}\omega_{uv}(Q_{uv}+\delta_{uv}Q_u)(\rho_{uv})_-$$

$$\leqslant \frac{\kappa\tau}{\tau-1}\left(\sum_{uv\in A}\eta_{uv}Q_{uv}\hat{c}_v+\sum_{uv\in A}Q_u\theta_{uv}(\hat{c}_v-\alpha_{uv})\right), \tag{4.16}$$

where the last line uses Lemma 4.10 and (4.9), to bound $\omega_{uv}(\rho_{uv})_- \leqslant \kappa\eta_{uv}\,(\hat{c}_v-\alpha_{uv})$.

Lemma 4.11. *It holds that*

$$\sum_{uv\in A}\eta_{uv}Q_{uv}\hat{c}_v \leqslant (\Delta_0(\mathcal{D})+\Delta_I(\hat{\mathcal{D}}))\,\langle c,Q\rangle_X\,.$$

Proof. Consider a decomposition of Q into flows on single source-sink paths. More precisely, let $\chi:\mathcal{P}_{\mathcal{D}}\to\mathbb{R}_+$ be so that

$$Q=\sum_{\gamma\in\mathcal{P}_{\mathcal{D}}}\chi(\gamma)\mathbb{1}_\gamma.$$

Note that the existence of such a decomposition is guaranteed by (2.1). Now we have

$$\sum_{uv\in A}\eta_{uv}Q_{uv}\hat{c}_v \leqslant \sum_{\gamma\in\mathcal{P}_{\mathcal{D}}}c_{\bar{\gamma}}Q_{\bar{\gamma}}\chi(\gamma)\sum_{uv\in\gamma}\eta_{uv} \leqslant (\Delta_0(\mathcal{D})+\Delta_I(\hat{\mathcal{D}}))\,\langle c,Q\rangle_X\,,$$

since for any $\gamma\in\mathcal{P}_{\mathcal{D}}$, we have

$$\sum_{uv\in\gamma}\eta_{uv}=|\gamma|+\log(1/\theta(\gamma))\leqslant\Delta_0(\mathcal{D})+\Delta_I(\hat{\mathcal{D}}).$$

□

It only remains to bound the latter term in (4.16). In order to do so, we would need the following result from [CL19].

Lemma 4.12 ([CL19, Lemma 2.11])**.** *For any $u\in V\setminus X$, it holds that*

$$\Psi_u(p)-\Psi_u(q)\leqslant\frac{2}{\kappa}(Q_u-P_u)_+\cdot\max_{v:uv\in A}\omega_{uv}+\sum_{v:uv\in A}(\hat{c}_v-\alpha_{uv})\left[Q_{uv}-\theta_{uv}Q_u\right]. \tag{4.17}$$

We omit a proof of the lemma as it is essentially identical to that of [CL19, Lem. 2.11]. In [CL19], for a fixed u, the probability distirbution specified by $\langle\theta_{uv}:uv\in A\rangle$ is uniform, but the argument works verbatim for any probability.

Lemma 4.13. *It holds that*

$$\frac{\tau-3}{\kappa\tau}\,\|(Q-P)_+\|_{\ell_1(\omega)}\leqslant(2\Delta_0(\mathcal{D})+\Delta_I(\hat{\mathcal{D}}))\langle c,Q\rangle_X+[\Psi(q)-\Psi(p)]\,.$$

Proof. Using Lemma 4.12 gives

$$\begin{aligned}\sum_{uv\in A}Q_u\theta_{uv}(\hat{c}_v-\alpha_{uv}) &\overset{(4.17)}{\leqslant}[\Psi(q)-\Psi(p)]+\frac{2}{\kappa\tau}\,\|(Q-P)_+\|_{\ell_1(\omega)}+\sum_{uv\in A}Q_{uv}\hat{c}_v\\ &\leqslant\ [\Psi(q)-\Psi(p)]+\frac{2}{\kappa\tau}\,\|(Q-P)_+\|_{\ell_1(\omega)}+\Delta_0(\mathcal{D})\langle c,Q\rangle_X.\end{aligned}$$

Combining this inequality with (4.16) and Lemma 4.11 gives

$$\kappa^{-1} \|(Q-P)_+\|_{\ell_1(\omega)} \leqslant \frac{\tau}{\tau-1}\left[\left(2\Delta_0(\mathcal{D}) + \Delta_I(\hat{\mathcal{D}})\right)\langle c, Q\rangle_X + (\Psi(q)-\Psi(p)) + \frac{2}{\kappa\tau}\|(Q-P)_+\|_{\ell_1(\omega)}\right],$$

completing the proof. □

Proof of Lemma 4.7. (4.11) follows from Lemma 4.13 and Lemma 4.9. To see that (4.12) follows from (4.11) and Lemma 4.13, use the fact that

$$\langle c, Q\rangle_X \leqslant \langle c, P\rangle_X + \frac{\|c\|_\infty}{\omega_{\min}}\|(Q-P)_+\|_{\ell_1(\omega)}\,. \qquad \square$$

Acknowledgements

We thank the anonymous referees for their helpful comments. This research was partially supported by NSF CCF-2007079 and a Simons Investigator Award.

References

[ABBS10] Jacob Abernethy, Peter Bartlett, Niv Buchbinder, and Isabelle Stanton. A regularization approach to metrical task systems. In *Algorithmic Learning Theory*, ALT 2010. Springer, 2010. 3

[Bar96] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 184–193, 1996. 2

[BBBT97] Yair Bartal, Avrim Blum, Carl Burch, and Andrew Tomkins. A polylog(n)-competitive algorithm for metrical task systems. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 711–719, New York, NY, USA, 1997. ACM. 2

[BBM06] Yair Bartal, Béla Bollobás, and Manor Mendel. Ramsey-type theorems for metric spaces with applications to online problems. *J. Comput. System Sci.*, 72(5):890–921, 2006. 2

[BCL+18] Sébastien Bubeck, Michael B. Cohen, Yin Tat Lee, James R. Lee, and Aleksander Madry. k-server via multiscale entropic regularization. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 3–16, 2018. 3

[BCLL21] Sébastien Bubeck, Michael B. Cohen, James R. Lee, and Yin Tat Lee. Metrical task systems on trees via mirror descent and unfair gluing. *SIAM J. Comput.*, 50(3):909–923, 2021. 2, 3, 4

[BCN14] Niv Buchbinder, Shahar Chen, and Joseph (Seffi) Naor. Competitive analysis via regularization. In *Proceedings of the Twenty-fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pages 436–444, Philadelphia, PA, USA, 2014. Society for Industrial and Applied Mathematics. 3

[BKRS00] Avrim Blum, Howard Karloff, Yuval Rabani, and Michael Saks. A decomposition theorem for task systems and bounds for randomized server problems. *SIAM J. Comput.*, 30(5):1624–1661, 2000. 2

[BLMN05] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type phenomena. *Ann. of Math. (2)*, 162(2):643–709, 2005. 2

[BLS92] Allan Borodin, Nathan Linial, and Michael E. Saks. An optimal on-line algorithm for metrical task system. *J. ACM*, 39(4):745–763, October 1992. 2

[CKR01] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 8–16, 2001. 10

[CL19] Christian Coester and James R. Lee. Pure entropic regularization for metrical task systems. In *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, pages 835–848, 2019. 3, 4, 5, 14, 15, 16, 17, 18, 19, 20, 21

[FM03] Amos Fiat and Manor Mendel. Better algorithms for unfair metrical task systems and applications. *SIAM Journal on Computing*, 32(6):1403–1422, 2003. 2

[FRT04] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004. 2

[Sei99] Steve Seiden. Unfair problems and randomized algorithms for metrical task systems. *Inf. Comput.*, 148(2):219–240, February 1999. 2