Dimension-Free Bounds for Chasing Convex Functions

C.J. Argue   CARGUE@ANDREW.CMU.EDU
Carnegie Mellon University, Pittsburgh, PA

Anupam Gupta   ANUPAMG@CS.CMU.EDU
Carnegie Mellon University, Pittsburgh, PA

Guru Guruganesh   GURUG@GOOGLE.COM
Google Research, Mountain View, CA

Abstract

We consider the problem of chasing convex functions, where functions arrive over time. The player takes actions after seeing the function, and the goal is to achieve a small function cost for these actions, as well as a small cost for moving between actions. While the general problem requires a polynomial dependence on the dimension, we show how to get dimension-independent bounds for well-behaved functions. In particular, we consider the case where the convex functions are κ-well-conditioned, and give an algorithm that achieves an $O(\sqrt{\kappa})$-competitiveness. Moreover, when the functions are supported on k-dimensional affine subspaces—e.g., when the function are the indicators of some affine subspaces—we get $O(\min(k, \sqrt{k \log T}))$-competitive algorithms for request sequences of length T. We also show some lower bounds, that well-conditioned functions require $\Omega(\kappa^{1/3})$-competitiveness, and k-dimensional functions require $\Omega(\sqrt{k})$-competitiveness.

Keywords: Online decision-making, online optimization, smooth online convex optimization, convex body/function chasing, subspace chasing, Steiner point.

1. Introduction

We consider the *convex function chasing* (CFC) problem defined by Friedman and Linial (1993), and independently studied under the name *smooth online convex optimization* (SOCO) by Lin et al. (2012, 2013). In this problem, an online player is faced with a sequence of convex functions over time, and has to choose a good sequence of responses to incur small function costs while also minimizing the movement cost for switching between actions. Formally, the player starts at some initial default action x_0, which is usually modeled as a point in $\mathbb{R}^d$. Convex functions $f_1, f_2, \ldots$ arrive online, one by one. Upon seeing the function $f_t : \mathbb{R}^d \to \mathbb{R}^+$, the player must choose an action x_t. The cost incurred by this action is

$$\|x_t - x_{t-1}\|_2 + f_t(x_t),$$

the former Euclidean distance term[1] being the *movement* or *switching* cost between the previous action x_{t-1} and the current action x_t, and the latter function value term being the *hit cost* at this new action x_t. Given some sequence of functions $\sigma = f_1, f_2, \ldots, f_T$, the online player's total cost

---

1. The problem can be defined for general metric spaces; in this paper we focus on the Euclidean norm.

for the associated sequence $X = (x_1, x_2, \ldots, x_T)$ is

$$\mathsf{cost}(X, \sigma) := \sum_{t=1}^{T} \Big(\|x_t - x_{t-1}\|_2 + f_t(x_t) \Big). \tag{1}$$

The *competitive ratio* for this player is $\max_\sigma \frac{\mathsf{cost}(ALG(\sigma), \sigma)}{\min_Y \mathsf{cost}(Y, \sigma)}$, the worst-case ratio of the cost of sequence of the player when given request sequence σ, to the cost of the optimal (dynamic) player for it (which is allowed to change its actions but has to also pay for its movement cost). The goal is to give an online algorithm that has a small competitive ratio.

The CFC/SOCO problem is usually studied in the setting where the action space is all of $\mathbb{R}^d$. We consider the generalized setting where the action space is any convex set $K \subseteq \mathbb{R}^d$. Formally, the set K is fixed before the arrival of f_1, and each action x_t must be chosen from K.

The CFC/SOCO problem captures many other problems arising in sequential decision making. For instance, it can be used to model problems in "right-sizing" data centers, charging electric cars, online logistic regression, speech animation, and control; see, e.g., works by Lin et al. (2012); Wang et al. (2014); Kim and Giannakis (2014); Goel et al. (2017, 2019) and the references therein. In all these problems, the action x_t of the player captures the state of the system (e.g., of a fleet of cars, or of machines in a datacenter), and there are costs associated both with taking actions at each timestep, and with changing actions between timesteps. The CFC/SOCO problem models the challenge of trading off these two costs against each other.

One special case of CFC/SOCO is the *convex body chasing* problem, where the convex functions are indicators of convex sets in $\mathbb{R}^d$. This special case itself captures the continuous versions of problems in online optimization that face similar tensions between taking near-optimal actions and minimizing movement: e.g., *metrical task systems* studied by Borodin et al. (1992); Bubeck et al. (2019a), *paging* and k-server (see (Bubeck et al., 2018; Buchbinder et al., 2019) for recent progress), and many others.

Given its broad expressive power, it is unsurprising that the competitiveness of CFC/SOCO depends on the dimension d of the space. Indeed, Friedman and Linial (1993) showed a lower bound of $\sqrt{d}$ on the competitive ratio for convex *body* chasing, and hence for CFC/SOCO as well. However, it was difficult to prove results about the upper bounds: Friedman and Linial gave a constant-competitive algorithm for *body* chasing for the case $d = 2$, and the *function* chasing problem was optimally solved for $d = 1$ by Bansal et al. (2015), but the general problem remained open for any higher dimensions. The logjam was broken in results by Bansal et al. (2018); Argue et al. (2019a) for some special cases, using ideas from convex optimization. After intense activity since then, algorithms with competitive ratio $O(\min(d, \sqrt{d \log T}))$ were given for the general CFC/SOCO problem by Argue et al. (2019b); Sellke (2019). These results qualitatively settle the question in the worst case—the competitive ratio is polynomial in d—although quantitative questions about the exponent for d remain.

However, this polynomial dependence on the dimension d can be very pessimistic, especially in cases when the convex functions have more structure. In these well-behaved settings, we may hope to get better results and thereby escape this curse of dimensionality. This motivates our work in this paper: we consider two such settings, and give dimension-independent guarantees for them.

Well-Conditioned Functions. The first setting we consider is when the functions f_t are all *well-conditioned convex* functions. Recall that a convex function has *condition number* κ if it is α-*strongly-convex* and β-*smooth* for some constants $\alpha, \beta > 0$ such that $\frac{\beta}{\alpha} = \kappa$. Moreover, we are

given a convex set K, and each point x_t we return must belong to K. (We call this the constrained CFC/SOCO problem; while constraints can normally be built into the convex functions, it may destroy the well-conditionedness in our setting, and hence we consider it separately.)

Our first main result is the following:

Theorem 1 (Upper Bound: Well-Conditioned Functions) *There is an $O(\sqrt{\kappa})$-competitive algorithm for constrained CFC/SOCO problem, where the functions have condition number at most κ.*

Observe that the competitiveness does not depend on d, the dimension of the space. Moreover, the functions can have very different coefficients of smoothness and strong convexity, as long as their ratio is bounded by κ. In fact, we give two algorithms. Our first algorithm is a direct generalization of the greedy-like *Move Towards Minimizer* algorithm of Bansal et al. (2015). While it only achieves a competitiveness of $O(\kappa)$, it is simpler and works for a more general class of functions (which we called "well-centered"), as well as for all ℓ_p norms. Our second algorithm is a constrained version of the *Online Balanced Descent* algorithm of Chen et al. (2018), and achieves the competitive ratio claimed in Theorem 1. We then show a lower bound in the same ballpark:

Theorem 2 (Lower Bound: Well-Conditioned Functions) *Any algorithm for chasing convex functions with condition number at most κ must have competitive ratio at least $\Omega(\kappa^{1/3})$.*

It remains an intriguing question to close the gap between the upper bound of $O(\sqrt{\kappa})$ from Theorem 1 and the lower bound of $\Omega(\kappa^{1/3})$ from Theorem 2. Since we show that $O(\kappa)$ and $O(\sqrt{\kappa})$ are respectively tight bounds on the competitiveness of the two algorithms mentioned above, closing the gap will require changing the algorithm.

Chasing Low-Dimensional Functions. The second case is when the functions are supported on low-dimensional subspaces of $\mathbb{R}^d$. One such special case is when the functions are indicators of k-dimensional affine subspaces; this problem is referred to as chasing subspaces. If $k = 0$ we are chasing points, and the problem becomes trivial. Friedman and Linial (1993) gave a constant-competitive algorithm for the first non-trivial case, that of $k = 1$ or *line chasing*. Antoniadis et al. (2016) simplified and improved this result, and also gave an $2^{O(d)}$-competitive algorithm for chasing general affine subspaces. Currently, the best bound even for 2-dimensional affine subspaces—i.e., planes—is $O(d)$, using the results for general CFC/SOCO.

Theorem 3 (Upper Bound: Low-Dimensional Chasing) *There is an $O(\min(k, \sqrt{k \log T}))$-competitive algorithm for chasing convex functions supported on affine subspaces of dimension at most k.*

The idea behind Theorem 3 is to perform a certain kind of dimension reduction: we show that any instance of chasing k-dimensional functions can be embedded into an $(2k + 1)$-dimensional instance, without changing the optimal solutions. Moreover, this embedding can be done online, and hence can be used to extend any $g(d)$-competitive algorithm for CFC/SOCO into a $g(2k + 1)$-competitive algorithm for k-dimensional functions.

1.1. Related Work

There has been prior work on dimension-independent bounds for other classes of convex functions. The Online Balanced Descent (OBD) algorithm of Chen et al. (2018) is α-competitive on Euclidean

metrics if each function f_t is α-locally-polyhedral (i.e., it grows at least linearly as we go away from the minimizer). Subsequent works of Goel and Wierman (2019); Goel et al. (2019) consider *squared Euclidean distances* and give algorithms with dimension-independent competitiveness of $\min(3 + O(1/\alpha), O(\sqrt{\alpha}))$ for α-strongly convex functions. The requirement of squared Euclidean distances in these latter works is crucial for their results: we show in Theorem 13 that no online algorithm can have dimension-independent competitiveness for *non-squared* Euclidean distances if the functions are only α-strongly convex (or only β-smooth). Observe that our algorithms do not depend on the actual value of the strong convexity coefficient α, only on the ratio between it and the smoothness coefficient β—so the functions f_t may have very different α_t, β_t values, and these α_t may even be arbitrarily close to zero.

A related problem is the notion of regret minimization, which considers the additive gap of the algorithm's cost (1) with respect to the best *static* action x^* instead of the multiplicative gap with respect to the best *dynamic* sequence of actions. The notions of competitive ratio and regret are known to be inherently in conflict: Andrew et al. (2013) showed that algorithms minimizing regret must have poor competitive ratios in the worst-case. Despite this negative result, many ideas do flow from one setting to the other. These is a vast body of work where the algorithm is allowed to move for free: see, e.g., books by Bubeck (2015); Hazan (2016); Shalev-Shwartz (2012) for many algorithmic ideas. This includes bounds comparing to the static optimum, and also to a dynamic optimum with a bounded movement cost Zinkevich (2003); Besbes et al. (2015); Mokhtari et al. (2016); Bubeck et al. (2019b).

Motivated by convergence and generalization bounds for learning algorithms, the path length of gradient methods have been studied by Oymak and Soltanolkotabi (2019); Gupta et al. (2019). Results for CFC/SOCO also imply path-length bounds by giving the same function repeatedly: the difference is that these papers focus on a specific algorithm (e.g., gradient flow/descent), whereas we design problem-specific algorithms (M2M or COBD).

The CFC/SOCO problem has been considered in the case with *predictions* or *lookahead*: e.g., when the next w functions are available to the algorithm. For example, Lin et al. (2012); Li et al. (2018) explore the value of predictions in the context of data-server management, and provide constant-competitive algorithms. For more recent work see, e.g., Lin et al. (2019) and the references therein.

1.2. Definitions and Notation

We consider settings where the convex functions f_t are non-negative and differentiable. Given constants $\alpha, \beta > 0$, a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is α-*strongly-convex* with respect to the norm $\| \cdot \|$ if for all $x, y \in \mathbb{R}^d$,

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \geq \frac{\alpha}{2} \|x - y\|^2,$$

and β-*smooth* if for all $x, y \in \mathbb{R}^d$,

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{\beta}{2} \|x - y\|^2.$$

A function f is κ-*well-conditioned* if there is a constant $\alpha > 0$ for which f is both α-strongly-convex and $\alpha\kappa$-smooth. Of course, we focus on the Euclidean ℓ_2 norm (except briefly in §D), and hence $\| \cdot \|$ denotes $\| \cdot \|_2$ unless otherwise specified.

In the following, we assume that all our functions f satisfy the *zero-minimum* property: i.e., that $\min_y f(y) = 0$. Else we can consider the function $g(x) = f(x) - \min_y f(y)$ instead: this is also non-negative valued, with the same smoothness and strong convexity as f. Moreover, the competitive ratio can only increase when we go from f to g, since the hit costs of both the algorithm and the optimum decrease by the same additive amount.

Cost. Consider a sequence $\sigma = f_1, \ldots, f_T$ of functions. If the algorithm moves from x_{t-1} to x_t upon seeing function f_t, the *hit cost* is $f_t(x_t)$, and the *movement cost* is $\|x_t - x_{t-1}\|$. Given a sequence $X = (x_1, \ldots, x_T)$ and a time t, define $\mathsf{cost}_t(X, \sigma) := \|x_t - x_{t-1}\| + f_t(x_t)$ to be the total cost (i.e., the sum of the hit and movement costs) incurred at time t. When the algorithm and request sequence σ are clear from context, let $X_{ALG} = (x_1, x_2, \ldots, x_T)$ denote the sequence of points that the algorithm plays on σ. Moreover, denote the offline optimal sequence of points by $Y_{OPT} = (y_1, y_2, \ldots, y_T)$. For brevity, we omit mention of σ and let $\mathsf{cost}_t(ALG) := \mathsf{cost}_t(X_{ALG}, \sigma)$ and $\mathsf{cost}_t(OPT) := \mathsf{cost}_t(Y_{OPT}, \sigma)$.

Potentials and Amortized Analysis. Given a potential Φ_t associated with time t, denote $\Delta_t \Phi := \Phi_t - \Phi_{t-1}$. Hence, for all the amortized analysis proofs in this paper, the goal is to show

$$\mathsf{cost}_t(ALG) + a \cdot \Delta_t \Phi \le b \cdot \mathsf{cost}_t(OPT)$$

for suitable parameters a and b. Indeed, summing this over all times gives

$$(\text{total cost of } ALG) + a(\Phi_T - \Phi_0) \le b \cdot (\text{total cost of } OPT).$$

Now if $\Phi_0 \le \Phi_T$, which is the case for all our potentials, we get that the cost of the algorithm is at most b times the optimal cost, and hence the algorithm is b-competitive.

Deterministic versus Randomized Algorithms. We only consider deterministic algorithms. This is without loss of generality by the observation in (Bansal et al., 2015, Theorem 2.1): given a randomized algorithm which plays the random point X_t at each time t, instead consider deterministically playing the "average" point $\mu_t := \mathbb{E}[X_t]$. This does not increase either the movement or the hit cost, due to Jensen's inequality and the convexity of the functions f_t and the norm $\|\cdot\|$.

2. Algorithms

We now give two algorithms for convex function chasing: §2.1 contains the simpler *Move Towards Minimizer* algorithm that achieves an $O(\kappa)$-competitiveness for κ-well-conditioned functions, and a more general class of *well-centered functions* (defined in Section 2.1.2). Then §2.2 contains the *Constrained Online Balanced Descent* algorithm that achieves the $O(\sqrt{\kappa})$-competitiveness claimed in Theorem 1.

2.1. Move Towards Minimizer: $O(\kappa)$-Competitiveness

The *Move Towards Minimizer* (M2M) algorithm was defined in Bansal et al. (2015).

> **The M2M Algorithm.** Suppose we are at position x_{t-1} and receive the function f_t. Let $x_t^* := \arg\min_x f_t(x)$ denote the minimizer of f_t. Consider the line segment with endpoints x_{t-1} and x_t^*, and let x_t be the unique point on this segment with $\|x_t - x_{t-1}\| = f_t(x_t) - f_t(x_t^*)$.[2] The point x_t is the one played by the algorithm.

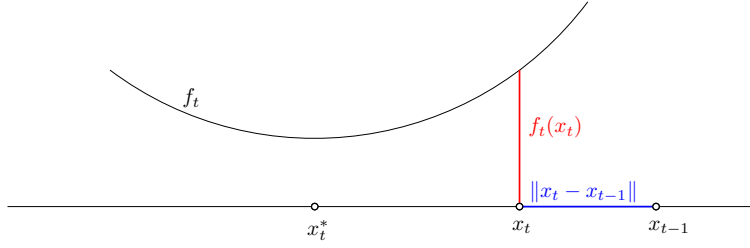2. Such a point is always unique when f_t is strictly convex.

Figure 1: The M2M *Algorithm in dimension $d = 1$.*

The intuition behind this algorithm is that one of two things happens: either the optimal algorithm OPT is at a point y_t near x_t^*, in which case we make progress by getting closer to OPT. Otherwise, the optimal algorithm is far away from x_t^*, in which case the hit cost of OPT is large relative to the hit cost of ALG.

As noted in §1.2, we assume that $f_t(x_t^*) = 0$, hence M2M plays a point x_t such that $\|x_t - x_{t-1}\| = f_t(x_t)$. Observe that the total cost incurred by the algorithm at time t is

$$\text{cost}_t(ALG) = f_t(x_t) + \|x_t - x_{t-1}\| = 2f_t(x_t) = 2\|x_t - x_{t-1}\|.$$

2.1.1. THE ANALYSIS

The proof of competitiveness for M2M is via a potential function argument. The potential function captures the distance between the algorithm's point x_t and the optimal point y_t. Specifically, fix an optimal solution playing the sequence of points $Y_{OPT} = (y_1, \ldots, y_T)$, and define

$$\Phi_t := \|x_t - y_t\|.$$

Observe that $\Phi_0 = 0$ and $\Phi_t \geq 0$.

Theorem 4 *With $c := 4 + 4\sqrt{2}$, for each t,*

$$\text{cost}_t(ALG) + 2\sqrt{2} \cdot \Delta_t \Phi \leq c \cdot \kappa \cdot \text{cost}_t(OPT). \tag{2}$$

Hence, the M2M *algorithm is $c\kappa$-competitive.*

The main technical work is in the following lemma, which will be used to establish the two cases in the analysis. Referring to Figure 2, imagine the minimizer for f_t as being at the origin, the point y as being the location of OPT, and the points x and γx as being the old and new position of ALG. Intuitively, this lemma says that either ALG's motion in the direction of the origin significantly reduces the potential, or OPT is far from the origin and hence has significant hit cost.

Lemma 5 (Structure Lemma) *Given any scalar $\gamma \in [0, 1]$ and any two vectors $x, y \in \mathbb{R}^d$, at least one of the following holds:*

(i) $\|y - \gamma x\| - \|y - x\| \leq -\frac{1}{\sqrt{2}}\|x - \gamma x\|.$

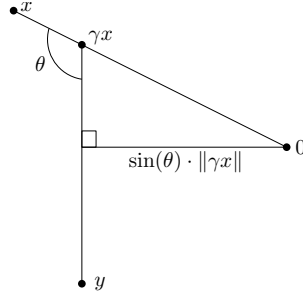(ii) $\|y\| \geq \frac{1}{\sqrt{2}}\|\gamma x\|.$

Figure 2: The Proof of Lemma 5.

Proof Let θ be the angle between x and $y - \gamma x$ as in Figure 2. If $\theta < \frac{\pi}{2}$, then $\|y\| \geq \|\gamma x\|$, and hence condition (ii) is satisfied. So let $\theta \in [\frac{\pi}{2}, \pi]$; using Figure 2 observe that

$$\|y\| \geq \sin(\theta) \cdot \|\gamma x\|. \tag{3}$$

Suppose condition (i) does not hold. Then

$$\|y - x\| < \tfrac{1}{\sqrt{2}} \|(1-\gamma)x\| + \|y - \gamma x\|.$$

Since both sides are non-negative, we can square to get

$$\|y - x\|^2 < \frac{1}{2}(1-\gamma)^2 \|x\|^2 + \sqrt{2} \cdot \|(1-\gamma)x\| \cdot \|y - \gamma x\| + \|y - \gamma x\|^2$$

$$\implies \|y - x\|^2 - \|y - \gamma x\|^2 < \frac{1}{2}(1-\gamma)^2 \|x\|^2 + \sqrt{2}(1-\gamma) \cdot \|x\| \cdot \|y - \gamma x\|.$$

The law of cosines gives

$$\|y - x\|^2 - \|y - \gamma x\|^2 = (1-\gamma)^2 \|x\|^2 - 2(1-\gamma)\cos(\theta) \cdot \|x\| \cdot \|y - \gamma x\|.$$

Substituting and simplifying,

$$\frac{1}{2}(1-\gamma)\|x\| < (\sqrt{2} + 2\cos(\theta))\|y - \gamma x\|.$$

As the LHS is non-negative, $\cos(\theta) > -\frac{1}{\sqrt{2}}$. Since $\theta \geq \frac{\pi}{2}$, it follows that $\sin(\theta) > \frac{1}{\sqrt{2}}$. Now, (3) implies that $\|y\| \geq \sin(\theta) \cdot \|\gamma x\| \geq \frac{1}{\sqrt{2}} \|\gamma x\|$. ∎

Proof [Proof of theorem 4] First, the change in potential can be bounded as

$$\Delta_t \Phi = \|x_t - y_t\| - \|x_{t-1} - y_{t-1}\| \leq \|x_t - y_t\| - \Big(\|x_{t-1} - y_t\| - \|y_t - y_{t-1}\| \Big).$$

The resulting term $\|y_t - y_{t-1}\|$ can be charged to the movement cost of OPT, and hence it suffices to show that

$$\mathsf{cost}_t(ALG) + 2\sqrt{2} \cdot \widetilde{\Delta}_t \Phi \leq (4 + 4\sqrt{2})\kappa \cdot f_t(y_t), \tag{4}$$

where $\widetilde{\Delta}_t \Phi := \|x_t - y_t\| - \|x_{t-1} - y_t\|$ denotes the change in potential due to the movement of ALG. Recall that x_t^* was the minimizer of the function f_t. The claim is translation invariant, so assume $x_t^* = 0$. This implies that $x_t = \gamma x_{t-1}$ for some $\gamma \in (0, 1)$. Lemma 5 applied to $y = y_t$, $x = x_{t-1}$ and γ, guarantees that one of the following holds:

7

(i) $\widetilde{\Delta}_t \Phi = \|x_t - y_t\| - \|x_{t-1} - y_t\| \leq -\frac{1}{\sqrt{2}}\|x_t - x_{t-1}\|.$

(ii) $\|y_t\| \geq \frac{1}{\sqrt{2}}\|x_t\|.$

Case I: Suppose $\widetilde{\Delta}_t \Phi \leq -\frac{1}{\sqrt{2}}\|x_t - x_{t-1}\|$. Since $\mathrm{cost}_t(ALG) \leq 2\|x_t - x_{t-1}\|$,

$$\mathrm{cost}_t(ALG) + 2\sqrt{2} \cdot \widetilde{\Delta}_t \Phi \leq 2\|x_t - x_{t-1}\| - 2\|x_t - x_{t-1}\| = 0$$
$$\leq (4 + 4\sqrt{2})\kappa \cdot f_t(y_t).$$

This proves (4).

Case II: Suppose that $\|y_t\| \geq \frac{1}{\sqrt{2}}\|x_t\|$. By the well-conditioned assumption on f_t (say, f_t is α_t-strongly-convex and $\alpha_t \kappa$ smooth) and the assumption that 0 is the minimizer of f_t, we have

$$f_t(x_t) \leq \frac{\alpha_t \kappa}{2}\|x_t\|^2 \leq \alpha_t \kappa\|y_t\|^2 \leq 2\kappa \cdot f_t(y_t). \tag{5}$$

By the triangle inequality and choice of x_t such that $f_t(x_t) = \|x_t - x_{t-1}\|$ we have

$$\widetilde{\Delta}_t \Phi = \|x_t - y_t\| - \|x_{t-1} - y_t\| \leq \|x_t - x_{t-1}\| = f_t(x_t).$$

Using $\mathrm{cost}_t(ALG) = 2f_t(x_t)$,

$$\mathrm{cost}_t(ALG) + 2\sqrt{2} \cdot \widetilde{\Delta}_t \Phi \leq 2f_t(x_t) + 2\sqrt{2}f_t(x_t)$$
$$\overset{(20)}{\leq} (4 + 4\sqrt{2})\kappa \cdot f_t(y_t).$$

This proves (4) and hence the bound (2) on the amortized cost. Now summing (2) over all times t, and using that $\Phi_t \geq 0 = \Phi_0$, proves the competitiveness. ∎

We extend Theorem 4 to the constrained setting (by a modified algorithm); see §C. We also extend the result to general norms by replacing Lemma 5 by Lemma 20; details appear in §D. Moreover, the analysis of M2M is tight: in Proposition 15 we show an instance for which the M2M algorithm has $\Omega(\kappa)$-competitiveness.

2.1.2. WELL-CENTERED FUNCTIONS

The proof of Theorem 4 did not require the full strength of the well-conditioned assumption. In fact, it only required that each function f_t is κ-well-conditioned "from the perspective of its minimizer x_t^*", namely that there is a constant α such that for all $x \in \mathbb{R}^d$,

$$\frac{\alpha}{2}\|x - x_t^*\|^2 \leq f_t(x) \leq \frac{\kappa\alpha}{2}\|x - x_t^*\|^2.$$

Motivated by this observation, we define a somewhat more general class of functions for which the M2M algorithm is competitive.

Definition 6 *Fix scalars $\kappa, \gamma \geq 1$. A convex function $f : \mathbb{R}^d \to \mathbb{R}^+$ with minimizer x^* is (κ, γ)-well-centered if there is a constant $\alpha > 0$ such that for all $x \in \mathbb{R}^d$,*

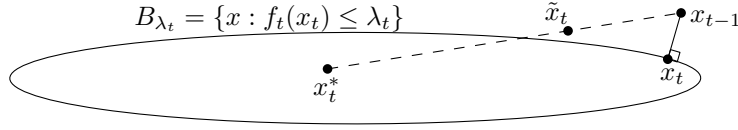$$\frac{\alpha}{2}\|x - x^*\|^\gamma \leq f(x) \leq \frac{\alpha\kappa}{2}\|x - x^*\|^\gamma.$$

Figure 3: The OBD *Algorithm and the comparison to* M2M. *The point x_{t-1} and the function f_t with minimizer x_t^* are given.* OBD *plays the point x_t and* M2M *plays the point $\tilde{x}_t$.*

We can now give a more general result.

Proposition 7 *If each function f_t is (κ, γ)-well centered, then with $c = 2 + 2\sqrt{2}$,*

$$\mathsf{cost}_t(ALG) + 2\sqrt{2} \cdot \Delta_t \Phi \leq c \cdot 2^{\gamma/2} \kappa \cdot \mathsf{cost}_t(OPT).$$

Hence, the M2M *algorithm is $c2^{\gamma/2}\kappa$-competitive.*

Proof Consider the proof of Theorem 4 and replace (20) by

$$f_t(x_t) \leq \frac{\alpha_t \kappa}{2} \|x_t\|^\gamma \leq \frac{\alpha_t \kappa}{2} \|y_t\|^\gamma \cdot 2^{\gamma/2} \leq 2^{\gamma/2} \kappa \cdot f_t(y_t).$$

The rest of the proof remains unchanged. ■

2.2. Constrained Online Balanced Descent: $O(\sqrt{\kappa})$-Competitiveness

The move-to-minimizer algorithm can pay a lot in one timestep if the function decreases slowly in the direction of the minimizer but decreases quickly in a different direction. In the unconstrained setting, the *Online Balanced Descent* algorithm addresses this by moving to a point x_t such that $\|x_t - x_{t-1}\| = f_t(x_t)$, except it chooses the point x_t to minimize $f_t(x_t)$. It therefore minimizes the instantaneous cost $\mathsf{cost}_t(ALG)$ among all algorithms that balance the movement and hit costs. This algorithm can be viewed geometrically as projecting the point x_{t-1} onto a level set of the function f_t; see Figure 3.

In the constrained setting, it may be the case that $\|x_t - x_{t-1}\| < f_t(x_t)$ for all feasible points. Accordingly, the *Constrained Online Balanced Descent* (COBD) algorithm moves to a point x_t that minimizes $f_t(x_t)$ subject to $\|x_t - x_{t-1}\| \leq f_t(x_t)$.

Formally, suppose that each f_t is α_t-strongly convex and $\beta_t := \kappa \alpha_t$-smooth, and let x_t^* be the (global) minimizer of f_t, which may lie outside K. As before, we assume that $f_t(x_t^*) = 0$.

> **The Constrained OBD Algorithm.** Let x_t be the solution to the (nonconvex) program $\min\{f_t(x) \mid \|x - x_{t-1}\| \leq f_t(x), x \in K\}$. Move to the point x_t. (Regarding efficient implementation of COBD, see Remark 18.)

As with M2M, the choice of x_t such that $\|x_t - x_{t-1}\| \leq f_t(x_t)$ implies that

$$\mathsf{cost}_t(ALG) = f_t(x_t) + \|x_t - x_{t-1}\| \leq 2f_t(x_t).$$

2.2.1. THE ANALYSIS.

Again, consider the potential function:

$$\Phi_t := \|x_t - y_t\|,$$

where x_t is the point controlled by the COBD algorithm, and y_t is the point controlled by the optimum algorithm. We first prove two useful lemmas. The first lemma is a general statement about β-smooth functions that is independent of the algorithm.

Lemma 8 *Let convex function f be β-smooth. Let x^* be the global minimizer of f, and suppose $f(x^*) = 0$ (as discussed in §1.2). Then for all $x \in \mathbb{R}^d$,*

$$\|\nabla f(x)\| \leq \sqrt{2\beta f(x)}.$$

Proof The proof follows (Bubeck, 2015, Lemma 3.5). Define $z := x - \frac{1}{\beta}\nabla f(x)$. Then

$$
\begin{aligned}
f(x) &\geq f(x) - f(z) && \text{(since } f(z) \geq 0) \\
&\geq \langle \nabla f(x), x - z \rangle - \frac{\beta}{2}\|x - z\|^2 && \text{(by } \beta\text{-smoothness)} \\
&= \langle \nabla f(x), \frac{1}{\beta}\nabla f(x) \rangle - \frac{1}{2\beta}\|\nabla f(x)\|^2 = \frac{1}{2\beta}\|\nabla f(x)\|^2.
\end{aligned}
$$

The conclusion follows. ■

The second lemma is specifically about COBD; its proof appears in Appendix B.

Lemma 9 *For each $t \geq 1$, there is a constant $\lambda \geq 0$ and a vector n in the normal cone to K at x_t such that $x_{t-1} - x_t = \lambda \nabla f_t(x_t) + n$.*

Theorem 10 *With $c = 2\sqrt{2\kappa}$, for each time t it holds that*

$$\text{cost}_t(ALG) + c \cdot \Delta_t \Phi \leq 2(2 + c) \cdot \text{cost}_t(OPT).$$

Hence, the COBD algorithm is $2(2 + c) = O(\sqrt{\kappa})$-competitive.

Proof As in the proof of Theorem 4, it suffices to show that

$$\text{cost}_t(ALG) + c \cdot \widetilde{\Delta}_t \Phi \leq 2(2 + c) \cdot f_t(y_t), \tag{6}$$

where $\widetilde{\Delta}_t \Phi := \|x_t - y_t\| - \|x_{t-1} - y_t\|$ is the change in potential due to the movement of ALG.

There are two cases, depending on the value of $f_t(y_t)$ versus the value of $f_t(x_t)$. In the first case, $f_t(y_t) \geq \frac{1}{2}f_t(x_t)$. The triangle inequality bounds $\widetilde{\Delta}_t \Phi = \|x_t - y_t\| - \|x_{t-1} - y_t\| \leq \|x_t - x_{t-1}\| \leq f_t(x_t)$. Also using $\text{cost}_t(ALG) \leq 2f_t(x_t)$, we have

$$\text{cost}_t(ALG) + c \cdot \widetilde{\Delta}_t \Phi \leq 2f_t(x_t) + cf_t(x_t) \leq 2(2 + c) \cdot f_t(y_t).$$

In the other case, $f_t(y_t) \leq \frac{1}{2}f_t(x_t)$. Note that this implies that x_t is not the minimizer of f_t on the set K. Any move in the direction of the minimizer gives a point in K with lower hit cost, but

10

this point cannot be feasible for the nonconvex program. Therefore, at the point x_t, the constraint relating the hit cost to the movement cost is satisfied with equality: $\|x_t - x_{t-1}\| = f_t(x_t)$.

Let θ_t be the angle formed by the vectors $\nabla f_t(x_t)$ and $y_t - x_t$; see Figure 4. We now have

$$
\begin{aligned}
-\langle \nabla f_t(x_t), y_t - x_t \rangle &\geq f_t(x_t) - f_t(y_t) + \frac{\alpha_t}{2}\|x_t - y_t\|^2 && \text{(by strong convexity)} \\
&\geq \frac{1}{2}f_t(x_t) + \frac{\alpha_t}{2}\|x_t - y_t\|^2 && \text{(since } f(y_t) \leq \tfrac{1}{2}f(x_t)) \\
\implies -\cos\theta_t &\geq \frac{\frac{1}{2}(f_t(x_t) + \alpha_t\|x_t - y_t\|^2)}{\|\nabla f_t(x_t)\| \cdot \|x_t - y_t\|} \\
&\geq \frac{\frac{1}{2}(f_t(x_t) + \alpha_t\|x_t - y_t\|^2)}{\sqrt{2\alpha_t \kappa\ f_t(x_t)} \cdot \|x_t - y_t\|} && \text{(by Lemma 8)} \\
&\geq \frac{1}{\sqrt{2\kappa}} && \text{(by the AM-GM inequality)}
\end{aligned}
$$

By Lemma 9, we have $x_{t-1} - x_t = \lambda \nabla f_t(x_t) + n$ for some n in the normal cone to K at point x_t. Since $y_t \in K$ we have $\langle n, y_t - x_t \rangle \leq 0$. This gives

$$
- \langle x_{t-1} - x_t, y_t - x_t \rangle = -\langle \lambda \nabla f_t(x_t) + n, y_t - x_t \rangle \geq -\lambda \nabla \langle f_t(x_t), y_t - x_t \rangle \tag{7}
$$

Furthermore, we have $\lambda \nabla f_t(x_t) = (x_{t-1} - x_t) - n$, and since $\langle x_{t-1} - x_t, n \rangle < 0$ we have

$$
\|x_{t-1} - x_t\| \leq \lambda \|\nabla f_t(x_t)\| \tag{8}
$$

Let φ_t be the angle formed by the vectors $x_{t-1} - x_t$ and $y_t - x_t$; see Figure 4.

Combining the previous three inequalities,

$$
\begin{aligned}
- \sec\varphi_t &= \frac{\|x_{t-1} - x_t\| \cdot \|y_t - x_t\|}{-\langle x_{t-1} - x_t, y_t - x_t \rangle} \\
&\leq \frac{\lambda\|\nabla f_t(x_t)\| \cdot \|y_t - x_t\|}{-\lambda\langle \nabla f_t, y_t - x_t \rangle} && \text{(by (7), (8))} \\
&= -\sec\theta_t \\
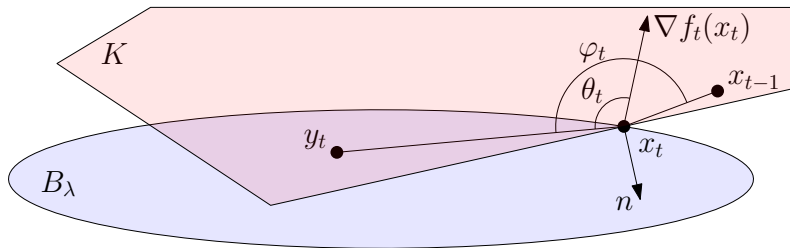&\leq \sqrt{2\kappa} = \frac{c}{2}
\end{aligned}
$$

Figure 4: Proof of Theorem 10, case when $f_t(y_t) \leq \frac{1}{2}f_t(x_t)$. B_λ is the sublevel set of f_t with x_t is on its boundary.

Now the law of cosines gives:

$$\|x_t - x_{t-1}\|^2 - 2\|x_t - x_{t-1}\| \cdot \|x_t - y_t\| \cos\varphi_t = \|x_{t-1} - y_t\|^2 - \|x_t - y_t\|^2.$$

Rearranging:

$$\|x_t - x_{t-1}\| = \left(\frac{\|x_{t-1} - y_t\| + \|x_t - y_t\|}{\|x_t - x_{t-1}\| - 2\|x_t - y_t\| \cos\varphi_t} \right) \left(\|x_{t-1} - y_t\| - \|x_t - y_t\| \right)$$

$$\leq \left(\frac{\|x_t - x_{t-1}\| + 2\|x_t - y_t\|}{\|x_t - x_{t-1}\| - 2\|x_t - y_t\| \cos\varphi_t} \right) \left(\|x_{t-1} - y_t\| - \|x_t - y_t\| \right)$$

$$\text{(triangle inequality)}$$

$$\leq -(\sec\varphi_t) \cdot \left(\|x_{t-1} - y_t\| - \|x_t - y_t\| \right).$$

To see the last inequality, recall that $-\cos\varphi_t > 0$; hence $\frac{a+b}{a+b(-\cos\varphi_t)} \leq \frac{a+b}{(a+b)(-\cos\varphi_t)} = -\sec\varphi_t$. Using that $\mathsf{cost}_t(ALG) = 2\|x_t - x_{t-1}\|$, we can rewrite the inequality above as

$$\mathsf{cost}_t(ALG) - (2\sec\varphi_t) \cdot \widetilde{\Delta}_t\Phi \leq 0.$$

Finally, observe that since $y_t \in B_{\lambda_t}$, we have $\widetilde{\Delta}_t\Phi \leq 0$. Using the fact that $-\sec(\varphi_t) \leq \frac{c}{2}$,

$$\mathsf{cost}_t(ALG) + c\widetilde{\Delta}_t\Phi \leq \mathsf{cost}_t(ALG) - (2\sec\varphi_t) \cdot \widetilde{\Delta}_t\Phi \leq 0 \leq 2(2+c) \cdot f_t(y_t).$$

This completes the proof. ∎

Again, our analysis of COBD is tight: In Proposition 17 we show an instance for which the COBD algorithm has $\Omega(\sqrt{\kappa})$-competitiveness, even in the unconstrained setting.

3. Chasing Low-Dimensional Functions

In this section we prove Theorem 3, our main result for chasing low-dimensional convex functions. We focus our attention to the case where the functions f_t are indicators of some affine subspaces K_t of dimension k, i.e., $f_t(x) = 0$ for $x \in K_t$ and $f_t(x) = \infty$ otherwise. (The extension to the case where we have general convex functions supported on k-dimensional affine subspaces follows the same arguments.) The main ingredient in the proof of chasing low-dimensional affine subspaces is the following dimension-reduction theorem,

Theorem 11 *Suppose there is a $g(d)$-competitive algorithm for chasing convex bodies in $\mathbb{R}^d$, for each $d \geq 1$. Then for any $k \leq d$, there is a $g(2k+1)$-competitive algorithm to solve instances of chasing convex bodies in $\mathbb{R}^d$ where each request lies in an affine subspace of dimension at most k.*

In particular, Theorem 11 implies that there is a $(2k+1)$-competitive algorithm for chasing subspaces of dimension at most k, and hence proves Theorem 3. The proof of Theorem 11 appears in Appendix B.

Using the results for CFC/SOCO, this immediately gives an $O(\min(k, \sqrt{k \log T}))$-competitive algorithm to chase convex bodies lying in k-dimensional affine subspaces. Moreover, the lower bound of Friedman and Linial (1993) immediately extends to show an $\Omega(\sqrt{k})$ lower bound for k-dimensional subspaces. Finally, the proof for k-dimensional functions follows the same argument, and is deferred for now.

References

Lachlan Andrew, Siddharth Barman, Katrina Ligett, Minghong Lin, Adam Meyerson, Alan Roytman, and Adam Wierman. A tale of two metrics: Simultaneous bounds on competitiveness and regret. In *Conference on Learning Theory*, pages 741–763, 2013.

Antonios Antoniadis, Neal Barcelo, Michael Nugent, Kirk Pruhs, Kevin Schewior, and Michele Scquizzato. Chasing convex bodies and functions. In *LATIN*, pages 68–81. Springer, Berlin, 2016. doi: 10.1007/978-3-662-49529-2_6. URL http://dx.doi.org/10.1007/978-3-662-49529-2_6.

C.J. Argue, Sébastien Bubeck, Michael B. Cohen, Anupam Gupta, and Yin Tat Lee. A nearly-linear bound for chasing nested convex bodies. In *SODA*, pages 117–122. SIAM, 2019a.

C.J. Argue, Anupam Gupta, Guru Guruganesh, and Ziye Tang. Chasing convex bodies with linear competitive ratio. *CoRR*, abs/1905.11877, 2019b. URL http://arxiv.org/abs/1905.11877.

Nikhil Bansal, Anupam Gupta, Ravishankar Krishnaswamy, Kirk Pruhs, Kevin Schewior, and Cliff Stein. A 2-competitive algorithm for online convex optimization with switching costs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

Nikhil Bansal, Martin Böhm, Marek Eliáš, Grigorios Koumoutsos, and Seeun William Umboh. Nested convex bodies are chaseable. *SODA*, 2018.

Omar Besbes, Yonatan Gur, and Assaf J. Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015. doi: 10.1287/opre.2015.1408. URL https://doi.org/10.1287/opre.2015.1408.

Allan Borodin, Nathan Linial, and Michael E. Saks. An optimal on-line algorithm for metrical task system. *J. Assoc. Comput. Mach.*, 39(4):745–763, 1992. ISSN 0004-5411. doi: 10.1145/146585.146588. URL http://dx.doi.org/10.1145/146585.146588.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–357, 2015. ISSN 1935-8237. doi: 10.1561/2200000050. URL http://dx.doi.org/10.1561/2200000050.

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. doi: 10.1145/3188745.3188798. URL https://doi.org/10.1145/3188745.3188798.

Sébastien Bubeck, Michael B. Cohen, James R. Lee, and Yin Tat Lee. Metrical task systems on trees via mirror descent and unfair gluing. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 89–97, 2019a. doi: 10.1137/1.9781611975482.6. URL https://doi.org/10.1137/1.9781611975482.6.

Sébastien Bubeck, Yuanzhi Li, Haipeng Luo, and Chen-Yu Wei. Improved path-length regret bounds for bandits. In *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, pages 508–528, 2019b. URL http://proceedings.mlr.press/v99/bubeck19b.html.

Niv Buchbinder, Anupam Gupta, Marco Molinaro, and Joseph (Seffi) Naor. k-servers with a smile: Online algorithms via projections. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 98–116, 2019. doi: 10.1137/1.9781611975482.7. URL https://doi.org/10.1137/1.9781611975482.7.

Niangjun Chen, Gautam Goel, and Adam Wierman. Smoothed online convex optimization in high dimensions via online balanced descent. *arXiv preprint arXiv:1803.10366*, 2018.

Joel Friedman and Nathan Linial. On convex body chasing. *Discrete Comput. Geom.*, 9(3):293–321, 1993. ISSN 0179-5376. doi: 10.1007/BF02189324. URL http://dx.doi.org/10.1007/BF02189324.

Gautam Goel and Adam Wierman. An online algorithm for smoothed regression and LQR control. *Proceedings of Machine Learning Research*, 89:2504–2513, 2019.

Gautam Goel, Niangjun Chen, and Adam Wierman. Thinking fast and slow: Optimization decomposition across timescales. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1291–1298. IEEE, 2017.

Gautam Goel, Yiheng Lin, Haoyuan Sun, and Adam Wierman. Beyond online balanced descent: An optimal algorithm for smoothed online optimization. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 1873–1883, 2019. URL http://papers.nips.cc/paper/8463-beyond-online-balanced-descent-an-optimal-algorithm-for-smoothed-online-o

Chirag Gupta, Sivaraman Balakrishnan, and Aaditya Ramdas. Path length bounds for gradient descent and flow. *CoRR*, abs/1908.01089, 2019. URL http://arxiv.org/abs/1908.01089.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016. ISSN 2167-3888. doi: 10.1561/2400000013.

Seung-Jun Kim and Geogios B Giannakis. Real-time electricity pricing for demand response using online convex optimization. In *ISGT 2014*, pages 1–5. IEEE, 2014.

Yingying Li, Guannan Qu, and Na Li. Online optimization with predictions and switching costs: Fast algorithms and the fundamental limit. *arXiv preprint arXiv:1801.07780*, 2018.

Minghong Lin, Zhenhua Liu, Adam Wierman, and Lachlan LH Andrew. Online algorithms for geographical load balancing. In *2012 international green computing conference (IGCC)*, pages 1–10. IEEE, 2012.

Minghong Lin, Adam Wierman, Lachlan L. H. Andrew, and Eno Thereska. Dynamic right-sizing for power-proportional data centers. *IEEE/ACM Trans. Netw.*, 21(5):1378–1391, 2013. doi: 10.1109/TNET.2012.2226216. URL http://dx.doi.org/10.1109/TNET.2012.2226216.

Yiheng Lin, Gautam Goel, and Adam Wierman. Online optimization with predictions and non-convex losses. *CoRR*, abs/1911.03827, 2019. URL http://arxiv.org/abs/1911.03827.

Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7195–7201. IEEE, 2016.

Samet Oymak and Mahdi Soltanolkotabi. Overparameterized nonlinear learning: Gradient descent takes the shortest path? In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pages 4951–4960, 2019. URL http://proceedings.mlr.press/v97/oymak19a.html.

Mark Sellke. Chasing convex bodies optimally. *CoRR*, abs/1905.11968, 2019. URL http://arxiv.org/abs/1905.11968.

Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012. ISSN 1935-8237. doi: 10.1561/2200000018. URL http://dx.doi.org/10.1561/2200000018.

Hao Wang, Jianwei Huang, Xiaojun Lin, and Hamed Mohsenian-Rad. Exploring smart grid and data center interactions for electric power load balancing. *ACM SIGMETRICS Performance Evaluation Review*, 41(3):89–94, 2014.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*, pages 928–936, 2003. URL http://www.aaai.org/Library/ICML/2003/icml03-120.php.

Appendix A. Lower Bounds

In this section, we show a lower bound of $\Omega(\kappa^{1/3})$ on the competitive ratio of convex function chasing for κ-well-conditioned functions. We also show that our analyses of the M2M and COBD algorithms are tight: that they have competitiveness $\Omega(\kappa)$ and $\Omega(\sqrt{\kappa})$ respectively. In both examples, we take $K = \mathbb{R}^d$ to be the action space.

A.1. A Lower Bound of $\Omega(\kappa^{1/3})$

The idea of the lower bound is similar to the $\Omega(\sqrt{d})$ lower bound Friedman and Linial (1993), which we now sketch. In this lower bound, the adversary eventually makes us move from the origin to some vertex $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_d)$ of the hypercube $\{-1, 1\}^d$. At time t, the request f_t forces us to move to the subspace $\{x \mid x_i = \varepsilon_i \ \forall i \leq t\}$. Not knowing the remaining coordinate values, it is best for us to move along the coordinate directions and hence incur the ℓ_1 distance of d. However

the optimal solution can move from the origin to ε along the diagonal and incur the ℓ_2 distance of $\sqrt{d}$. Since the functions f_t in this example are not well-conditioned, we approximate them by well-conditioned functions; however, this causes the candidate OPT to also incur nonzero hit costs, leading to the lower bound of $\Omega(\kappa^{1/3})$ when we balance the hit and movement costs.

We begin with a lemma analyzing a general instance defined by several parameters, and then achieve multiple lower bounds by appropriate choice of the parameters.

Lemma 12 *Fix a dimension d, constants $\gamma > 0$ and $\lambda \geq \mu \geq 0$. Given any algorithm ALG for chasing convex functions, there is a request sequence $f_1, f_2, \ldots, f_d$ that satisfies:*

(i) *Each f_t is 2μ-strongly-convex and 2λ-smooth (hence (λ/μ)-well-conditioned.)*

(ii) $OPT \leq \gamma(1 + \mu d^{3/2}\gamma)\sqrt{d}.$

(iii) $ALG \geq (\gamma - \frac{1}{4\lambda})d.$

Proof Consider the instance where at each time $t \in \{1, \ldots, d\}$, we pick a uniformly random value $\varepsilon_t \in \{-1, 1\}$, and set

$$f_t(x) = \lambda \sum_{i=1}^{t}(x_i - \gamma\varepsilon_i)^2 + \mu \sum_{i=t+1}^{d} x_i^2.$$

One candidate for OPT is to move to the point $\gamma\boldsymbol{\varepsilon} := (\gamma\varepsilon_1, \gamma\varepsilon_2, \ldots, \gamma\varepsilon_d)$, and take all the functions at that point. The initial movement costs $\gamma\sqrt{d}$, and the tth timestep costs $f_t(\gamma\boldsymbol{\varepsilon}) = \mu(d - t)\gamma^2$. Hence, the total cost over the sequence is at most

$$\gamma\sqrt{d} + \mu\binom{d}{2}\gamma^2 \leq \gamma\left(1 + \mu d^{3/2}\gamma\right)\sqrt{d}.$$

Suppose the algorithm is at the point $\mathbf{z} = (z_1, \ldots, z_d)$ after timestep $t - 1$, and it moves to point $\mathbf{z}' = (z_1', \ldots, z_d')$ at the next timestep. Moreover, suppose the algorithm sets $z_t' = a$ when it sees $\varepsilon_t = 1$, and sets $z_t' = b$ if $\varepsilon_t = -1$. Then for timestep t, the algorithm pays in expectation at least

$$\frac{1}{2}[\lambda(a - \gamma)^2 + |a - z_t|] + \frac{1}{2}[\lambda(b + \gamma)^2 + |b - z_t|]$$

$$= \frac{\lambda}{2}\left[(a^2 - 2\gamma a + \gamma^2) + (b^2 + 2\gamma b + \gamma^2)\right] + \frac{1}{2}[|a - z_t| + |b - z_t|]$$

$$\geq \frac{\lambda}{2}\left[(a^2 - 2\gamma a + \gamma^2) + (b^2 + 2\gamma b + \gamma^2)\right] + \frac{1}{2}(a - b)$$

$$= \frac{\lambda}{2}\left[\left(a^2 - \left(2\gamma - \frac{1}{\lambda}\right)a + \gamma^2\right) + \left(b^2 + \left(2\gamma - \frac{1}{\lambda}\right)b + \gamma^2\right)\right]$$

$$\geq \gamma - \frac{1}{4\lambda}.$$

The last inequality follows from choosing $a = \gamma - 1/(2\lambda)$ and $b = \gamma + 1/(2\lambda)$ to minimize the respective quadratics. Hence, in expectation, the algorithm pays at least $\gamma - \frac{1}{4\lambda}$ at each time t. Summing over all times, we get a lower bound of $(\gamma - \frac{1}{4\lambda})d$ on the algorithm's cost. ∎

In particular, Lemma 12 implies a competitive ratio of at least

$$\left(\frac{\gamma - 1/(4\lambda)}{\gamma(1 + \mu d^{3/2}\gamma)}\right)\sqrt{d}$$

for chasing a class of functions that includes $f_1, \ldots, f_d$. It is now a simple exercise in choosing constants to get a lower bound on the competitiveness of any algorithm for chasing κ-well-conditioned functions, α-strongly-convex functions, and β-smooth functions.

Proposition 13 *The competitive ratio of any algorithm for chasing convex functions with condition number κ is $\Omega(\kappa^{1/3})$. Moreover, the competitive ratio of any algorithm for chasing α-strongly-convex (resp., β-smooth) functions is $\Omega(\sqrt{d})$.*

Proof For κ-strongly convex functions, apply Lemma 12 with dimension $d = \kappa^{2/3}$, constants $\gamma = \lambda = 1$ and $\mu = \kappa^{-1} = d^{-3/2}$. This shows a gap of $\Omega(\sqrt{d}) = \Omega(\kappa^{1/3})$. For α-strongly convex functions, choose $\mu = \alpha/2$, $\gamma = 1/(d^{3/2}\alpha)$, and $\lambda = 1/\gamma = d^{3/2}\alpha$. Finally, for β-smooth functions, choose $\lambda = \beta/2$, $\gamma = 1/\beta$, and $\mu = 0$. ∎

A.2. A Lower Bound Example for M2M

We show that the M2M algorithm is $\Omega(\kappa)$-competitive, even in $\mathbb{R}^2$. The essential step of the proof is the following lemma, which shows that, in a given timestep, ALG can be forced to pay $\Omega(k)$ times as much as some algorithm $Y = (y_1, \ldots, y_t)$ (we think of Y as a candidate for OPT) while at each step t, ALG does not move any closer to y_t.

Lemma 14 *Fix $\kappa > 0$. Suppose that $(x_1, \ldots, x_{t-1})$ is defined by the M2M algorithm and $Y = (y_1, \ldots, y_{t-1})$ is a point sequence such that $y_{t-1} \neq x_{t-1}$. Define the potential*

$$\Phi_s = \|x_s - y_s\|.$$

Then there is a κ-well-conditioned function f_t and a choice of y_t such that

(i) $\mathsf{cost}_t(ALG) \geq \Omega(1) \cdot \Phi_{t-1} \geq \Omega(k) \cdot \mathsf{cost}_t(Y)$

(ii) $\Phi_t \geq \Phi_{t-1}$, *and hence $y_t \neq x_t$.*

Proof Observe that if we modify an instance by an isometry the algorithm's sequence will also change by the same isometry. So we may assume that $x_{t-1} = (\gamma, \gamma)$ and $y_{t-1} = (2\gamma, 0)$, for some $\gamma > 0$. (See Figure 5.) Define

$$f_t(x) = \frac{1}{4\gamma}\left(\frac{1}{\kappa} \cdot x_1^2 + x_2^2\right).$$

Note that f_t is κ-well-conditioned. It is easily checked that $x_t = \lambda x_{t-1}$ for some $\lambda > \frac{1}{2}$ (recall that x_t is chosen to satisfy $f_t(x_t) = \|x_t - x_{t-1}\|$). Thus ALG pays:

$$\mathsf{cost}_t(ALG) = 2f_t(x_t) = 2\lambda^2 \cdot f_t(x_{t-1}) \geq \Omega(1) \cdot \gamma = \Omega(1) \cdot \Phi_{t-1}. \tag{9}$$

We choose $y_t = y_{t-1}$ so that the cost of Y is:

$$\mathsf{cost}_t(Y) = f_t(y_t) \leq O\left(\frac{1}{\kappa}\right) \cdot \gamma = O\left(\frac{1}{\kappa}\right) \cdot \Phi_{t-1}. \tag{10}$$

Multiplying (10) by $\Omega(\kappa)$ and combining with (9) completes the proof of *(i)*. The statement in *(ii)* follows from the fact that x_t, x_{t-1} and y_t form a right triangle with leg Φ_{t-1} and hypotenuse Φ_t. ∎
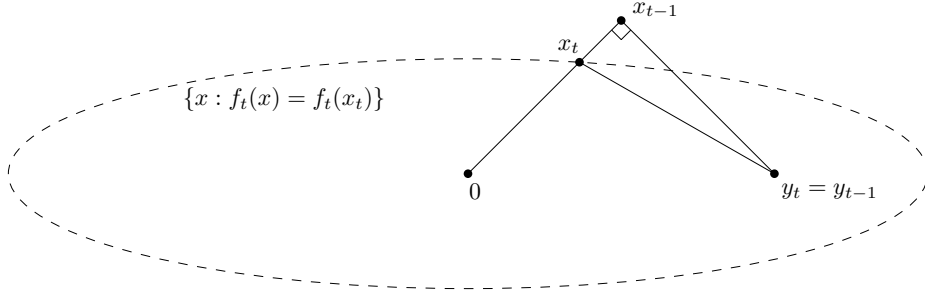
Figure 5: Proof of Lemma 14 showing that M2M *is* $\Omega(\kappa)$-*competitive.*

Proposition 15 *The* M2M *algorithm is* $\Omega(\kappa)$ *competitive for chasing* κ-*well-conditioned functions.*

Proof Suppose that before the first timestep, y_0 moves to e_1 and incurs cost 1. Now consider the instance given by repeatedly applying Lemma 14 for T timesteps. For each time t, we have $\Phi_t \geq \Phi_0$. Thus,

$$\text{cost}_t(ALG) \geq \Omega(1) \cdot \Phi_{t-1} \geq \Omega(1) \cdot \Phi_0 = \Omega(1).$$

Summing over all time, ALG pays $\text{cost}(ALG) \geq \Omega(T)$. Meanwhile, our candidate OPT has paid at most $O(\frac{1}{\kappa}) \cdot \text{cost}(ALG) + 1$. The proof is completed by choosing $T \geq \Omega(\kappa)$. ∎

A.3. A Lower Bound Example for COBD

We now give a lower bound for the COBD algorithm.[3] In the proof of Proposition 10 we showed that the angle θ_t between $y_t - x_t$ and $x_{t-1} - x_t$ satisfies $-\sec(\theta_t) \leq O(\sqrt{\kappa})$. This bound corresponds directly determines to the competitiveness of COBD. The essence of the lower bound is to give an example where $-\sec(\theta_t) \geq \Omega(\sqrt{\kappa})$.

Much like M2M, the key to showing that COBD is $\Omega(\sqrt{\kappa})$-competitive lies in constructing a single "bad timestep" that can be repeated until it dominates the competitive ratio. In the case of COBD, this timestep allows us to convert the potential into cost to ALG at a rate of $\Omega(\sqrt{\kappa})$.

Lemma 16 *Fix* $\kappa \geq 1$. *Suppose that* x_t *is defined by the* COBD *algorithm and that* $Y = (y_1, \ldots, y_{t-1})$ *is a point sequence such that* $y_{t-1} \neq x_{t-1}$. *Define the potential*

$$\Phi_s = \|x_s - y_s\|.$$

Then there is a κ-*well-conditioned function* f_t *and a choice of* y_t *such that*

   (i) $\text{cost}_t(ALG) \geq \Omega(\frac{1}{\sqrt{\kappa}})\Phi_{t-1}.$

   (ii) $\text{cost}_t(ALG) \geq \Omega(\sqrt{\kappa})(-\Delta_t\Phi).$

   (iii) $\text{cost}_t(Y) = 0.$

---

3. The lower bound example is valid even in the unconstrained setting, where COBD and OBD are the same algorithm.

Proof Observe that modifying an instance by an isometry will modify the algorithm's sequence by the same isometry. After applying an appropriate isometry, we will define

$$f_t(x) = \alpha(x_1^2 + \kappa x_2^2)$$

for some $\alpha > 0$ to be chosen later and $y_t = y_{t-1}$. We claim that this can be done such that:

(a) $y_t = y_{t-1} = 0$,

(b) $\|x_t - x_{t-1}\| = \frac{1}{2\sqrt{\kappa}}\|x_{t-1} - y_{t-1}\|$ (which in turn is equal to $\frac{1}{2\sqrt{\kappa}}\|x_{t-1}\|$).

(c) $x_t = \gamma \begin{bmatrix} \sqrt{\kappa} \\ 1 \end{bmatrix}$ for some $\gamma > 0$,

For any $\alpha > 0$, there is point a x_α on the ray $\left\{ \gamma \begin{bmatrix} \sqrt{\kappa} \\ 1 \end{bmatrix} : \gamma > 0 \right\}$ such that $f_t(x_\alpha) = \frac{1}{2\sqrt{\kappa}}\|x_{t-1} - y_{t-1}\|$. Let

$$x_\alpha^- := x_\alpha + \left(\frac{1}{2\sqrt{\kappa}}\|x_{t-1} - y_{t-1}\| \right) \frac{\nabla f_t(x_\alpha)}{\|\nabla f_t(x_\alpha)\|}.$$

Note that x_α^- is defined so that applying COBD to x_α^- and f_t outputs the point x_α. Then $\|x_\alpha^-\|$ increases continuously from $\frac{1}{2\sqrt{\kappa}}\|x_{t-1} - y_{t-1}\|$ to ∞ as α ranges from 0 to ∞. Choose α such that $\|x_\alpha^-\| = \|x_{t-1} - y_{t-1}\|$, and pick the isometry that maps y_{t-1} to 0 and x_{t-1} to x_α^-. The claim follows.

Now, (a) and (b) imply that

$$\mathrm{cost}_t(ALG) = 2\|x_t - x_{t-1}\| = \frac{1}{\sqrt{\kappa}}\|x_{t-1}\| = \frac{1}{\sqrt{\kappa}}\Phi_{t-1}.$$

This proves (i). Furthermore, (b) and the triangle inequality give

$$\|x_t\| \geq \|x_{t-1}\| - \|x_{t-1} - x_t\| = (2\sqrt{\kappa} - 1) \cdot \|x_{t-1} - x_t\| \geq \sqrt{\kappa} \cdot \|x_{t-1} - x_t\|. \tag{11}$$

There are $\eta, \nu > 0$[4] such that

$$x_{t-1} - x_t = \eta \nabla f_t(x_t) = \nu \begin{bmatrix} 1 \\ \sqrt{\kappa} \end{bmatrix}.$$

Letting θ_t be the angle between $x_{t-1} - x_t$ and $y_t - x_t = -x_t$ (cf. Figure 4) we have

$$-\cos(\theta_t) = -\frac{\langle x_{t-1} - x_t, -x_t \rangle}{\|x_{t-1} - x_t\| \cdot \| - x_t\|} = \frac{2\sqrt{\kappa}}{1 + \kappa} \leq \frac{2}{\sqrt{\kappa}}. \tag{12}$$

We now mirror the argument used in the proof of Theorem 10 relating $\mathrm{cost}_t(ALG)$ to $\cos(\theta_t)$.

$$\begin{aligned} \mathrm{cost}_t(ALG) &= 2\|x_t - x_{t-1}\| \\ &= \frac{\|x_{t-1}\| + \|x_t\|}{\|x_{t-1} - x_t\| - 2\|x_t\|\cos(\theta_t)} \cdot (-\Delta_t \Phi) \qquad \text{(Law of Cosines, substitution)} \end{aligned}$$

---

4. We omit the exact values (which depend on κ and $\|x_{t-1}\|$) as ν cancels out in the next step.

$$\geq \frac{\|x_t\|}{(1/\sqrt{\kappa})\|x_t\| + (4/\sqrt{\kappa}) \cdot \|x_t\|} \cdot (-\Delta_t \Phi) \qquad \text{(by (11) and (12))}$$

$$= \frac{\sqrt{\kappa}}{5}(-\Delta_t \Phi).$$

Finally, observe that $\text{cost}_t(Y) = f_t(0) = 0$. ∎

We can now get a lower bound on the competitiveness of COBD.

Proposition 17 COBD *is $\Omega(\sqrt{\kappa})$ competitive for chasing κ-well-conditioned functions.*

Proof Suppose that before the first timestep, y_0 moves to e_1 and incurs cost 1. Now consider the instance given by repeatedly applying Lemma 16 for T timesteps. $\text{cost}(OPT) = 1$, so it remains to show that $\text{cost}(ALG) = \Omega(\sqrt{k})$. Let $\Phi_{min} := \min\{\Phi_1, \ldots, \Phi_T\}$. Using (i) and summing over all time we have

$$\text{cost}(ALG) \geq \frac{1}{\sqrt{\kappa}} \sum_{t=0}^{T-1} \Phi_t \geq \frac{T}{\sqrt{\kappa}} \Phi_{min}. \qquad (13)$$

Using (ii) and summing over all time (and using that ALG incurs nonnegative cost at each step),

$$\text{cost}(ALG) \geq \Omega(\sqrt{\kappa})(\Phi_0 - \Phi_{min}) = \Omega(\sqrt{\kappa})(1 - \Phi_{min}) \qquad (14)$$

If $\Phi_{min} \geq \frac{1}{2}$ then $\text{cost}(ALG) \geq \frac{T}{2\sqrt{k}}$ by (13), else $\Phi_{min} < \frac{1}{2}$ and we have $\text{cost}(ALG) \geq \Omega(\sqrt{k})$ by (14). Choosing $T = \kappa$ completes the proof. ∎

Appendix B. Proofs from § 2, 3

Proof [Proof of Lemma 9] Let $r = \|x_t - x_{t-1}\|$. We claim that x_t is the solution to the following *convex* program:

$$\begin{aligned} \min \quad & f_t(x) \\ \text{s.t.} \quad & \|x - x_{t-1}\|^2 \leq r^2 \\ & x \in K \end{aligned}$$

Given this claim, the KKT conditions imply that there is a constant $\gamma \geq 0$ such that $\nabla f_t(x) + \gamma(x_t - x_{t-1})$ is in the normal cone to K at x_t and the result follows.

We now prove the claim. Assume for a contradiction that the solution to this program is a point $z \neq x_t$. We have $f_t(z) < f_t(x_t)$. Since $z \in K$ and x_t is the optimal solution to the nonconvex program $\min\{f_t(x) \mid \|x - x_{t-1}\| \leq f_t(x), x \in K\}$, we have $f(z) < \|z - x_{t-1}\|$. But considering the line segment with endpoints z and x_{t-1}, the intermediate value theorem implies that there is a point z' on this segment such that $f(z') = \|z' - x_{t-1}\|$. This point z' is feasible for the nonconvex program and

$$f(z') = \|z' - x_{t-1}\| < \|z - x_{t-1}\| = f(z) < f(x).$$

This contradicts the choice of x_t. The claim is proven, hence the proof of the lemma is complete. ∎

Remark 18 *The convex program given in the proof can be used to to find x_t efficiently. In particular, let r^* denote the optimal value to the nonconvex program. For a given r, if the solution to the convex program satisfies $f_t(x) < r$, then $r^* < r$. Otherwise, $r^* \geq r$. Noting that $0 \leq f_t(x_t) \leq f_t(x_{t-1})$, run a binary search to find r^* beginning with $r = \frac{1}{2} f_t(x_{t-1})$.*

Proof [Proof of Theorem 11] Suppose we have a chasing convex bodies instance $K_1, K_2, \ldots, K_T$ such that each K_t lies in some k-dimensional affine subspace. We construct another sequence $K'_1, \ldots, K'_T$ such that (a) there is a single $2k + 1$ dimensional linear subspace L that contains each K'_t, and (b) there is a feasible point sequence $x_1, \ldots, x_T$ of cost C for the initial instance if and only if there is a feasible point sequence $x'_1, \ldots, x'_T$ for the transformed instance with the same cost. We also show that the transformation from K_t to K'_t, and from x'_t back to x_t can be done online, resulting in the claimed algorithm.

Let $\mathrm{span}(S)$ denote the affine span of the set $S \subseteq \mathbb{R}^d$. Let $\dim(A)$ denote the dimension of an affine subspace $A \subseteq \mathbb{R}^d$. The construction is as follows: let L be an arbitrary $(2k + 1)$-dimensional linear subspace of $\mathbb{R}^d$ that contains K_1. We construct online a sequence of affine isometries $R_1, \ldots, R_T$ such that for each $t > 1$:

   (i) $R_t(K_t) \subseteq L$.

   (ii) $\|R_t(x_t) - R_{t-1}(x_{t-1})\| = \|x_t - x_{t-1}\|$ for any $x_{t-1} \in K_{t-1}$ and $x_t \in K_t$.

Setting $x'_t = R_t(x_t)$ then achieves the goals listed above. To get the affine isometry R_t we proceed inductively: let R_1 be the identity map, and suppose we have constructed R_{t-1}. Let $A_t := \mathrm{span}(R_{t-1}(K_t) \cup R_{t-1}(K_{t-1}))$. Note that $\dim(A_t) \leq 2k + 1$. Let ρ_t be an affine isometry that fixes $\mathrm{span}(R_{t-1}(K_{t-1}))$ and maps $\mathrm{span}(R_{t-1}(K_t))$ into L. Now define $R_t = \rho_t \circ R_{t-1}$. Property (i) holds by construction. Moreover, since $x_{t-1} \in K_{t-1}$, we have $R_t(x_{t-1}) = R_{t-1}(x_{t-1})$. Furthermore, R_t is an isometry and hence preserves distances. Thus,

$$\|R_t(x_t) - R_{t-1}(x_{t-1})\| = \|R_t(x_t) - R_t(x_{t-1})\| = \|x_t - x_{t-1}\|.$$

This proves (ii).

Note that $R(x_1, \ldots, x_T) := (R_1(x_1), \ldots, R_T(x_T))$ is a cost-preserving bijection between point sequences that are feasible for $\{K_t\}_t$ and $\{K'_t\}_t$ respectively. It now follows that the instances $\{K_t\}_t$ and $\{K'_t\}_t$ are equivalent in the sense that $OPT(K'_1, \ldots, K'_T) = OPT(K_1, \ldots, K_T)$, and an algorithm that plays points $x'_t \in K'_t$ can be converted into an algorithm of equal cost that plays points $x_t \in K_t$ by letting $x_t = R_t^{-1}(x'_t)$. However, each of $K'_1, \ldots, K'_T$ is contained in the $(2k+1)$ dimensional subspace L, and thus we get the $g(2k + 1)$-competitive algorithm. ∎

Appendix C. Constrained M2M

We give a generalized version of the M2M algorithm for the constrained setting where the action space $K \subseteq \mathbb{R}^d$ is an arbitrary convex set. This algorithm achieves the same $O(\sqrt{\kappa})$-competitiveness respectively as in the unconstrained setting.

The idea is to move towards $x^*_{K,t}$, the minimizer of f_t among feasible points, rather than the global minimizer. The proof of the algorithm's competitiveness proceeds similarly to the proof in the unconstrained setting. The difference is that it takes more care to show that $f(x_t) \leq O(\kappa)f(y_t)$ in Case II.

The Constrained M2M Algorithm. Suppose we are at position x_{t-1} and receive the function f_t. Let $x_{K,t}^* := \arg\min_{x \in K} f_t(x)$ denote the minimizer of f_t among points in K. Consider the line segment with endpoints x_{t-1} and $x_{K,t}^*$, and let x_t be the unique point on this segment with $\|x_t - x_{t-1}\| = f_t(x_t) - f_t(x_{K,t}^*)$.[5] The point x_t is the one played by the algorithm.

Note that we assume that the global minimum value of f_t is 0, as before. However, the minimum value of f_t on the action space K could be strictly positive.

Proposition 19 *With $c = 25(2 + 2\sqrt{2})$, for each t,*

$$\mathsf{cost}_t(ALG) + 2\sqrt{2} \cdot \Delta_t \Phi \leq c \cdot \kappa \cdot \mathsf{cost}_t(OPT). \tag{15}$$

Hence, the constrained M2M *algorithm is $c\kappa$-competitive.*

Proof As in the proof of Theorem 4, we begin by applying the structure lemma. This time, we use $x_{K,t}^*$ to be the origin. The proof of Case I is identical.

Case II: Suppose that $\|y_t - x_{K,t}^*\| \geq \frac{1}{\sqrt{2}}\|x_t - x_{K,t}^*\|$. Let $x_t^* := \arg\min_x f_t(x)$ denote the global minimizer of f_t. As before, we assume $f_t(x_t^*) = 0$, and we translate such that $x_t^* = 0$.

We now show that $f_t(x_t) \leq 25\kappa f_t(y_t)$. If $f_t(x_t) \leq 25\kappa f_t(x_{K,t}^*)$, then since $f(y_t) \geq f_t(x_{K,t}^*)$, we are done. So suppose that $f_t(x_t) > 25\kappa f_t(x_{K,t}^*)$. Now strong convexity and smoothness imply

$$\|x_t\|^2 \geq \frac{2}{\kappa\alpha_t} f_t(x_t) \geq 25 \cdot \frac{2}{\alpha_t} f(x_{K,t}^*) \geq 25\|x_{K,t}^*\|^2. \tag{16}$$

Thus $\|x_t\| \geq 5\|x_{K,t}^*\|$. One application of the triangle inequality gives $\|x_t - x_{K,t}^*\| \geq \|x_t\| - \|x_{K,t}^*\| \geq 4\|x_{K,t}^*\|$. Using the triangle inequality again, we get

$$\|x_t\| \leq \|x_t - x_{K,t}^*\| + \|x_{K,t}^*\| \leq \frac{5}{4}\|x - x_{K,t}^*\|, \tag{17}$$

and

$$\|y_t\| \geq \|y - x_{K,t}^*\| - \|x_{K,t}^*\| \geq \left(\frac{1}{\sqrt{2}} - \frac{1}{4}\right)\|x - x_{K,t}^*\| \geq \frac{1}{4}\|x - x_{K,t}^*\| \tag{18}$$

Combining these two, we have

$$\|y_t\| \geq \|x_t\| \geq \frac{1}{5}\|x_t\| \tag{19}$$

Finally, we have

$$f_t(x_t) \leq \frac{\alpha_t\kappa}{2}\|x_t\|^2 \leq \frac{5\alpha_t\kappa}{2}\|y_t\|^2 \leq 25\kappa \cdot f_t(y_t). \tag{20}$$

We now proceed as in the proof of Theorem 4. ∎

---

5. Such a point is always unique when f_t is strictly convex.

Appendix D. A Structure Lemma for General Norms

We can extend the $O(\kappa)$-competitiveness guarantee for M2M for all norms, by replacing Lemma 5 by the following Lemma 20 in Theorem 4, and changing some of the constants in the latter accordingly.

Lemma 20 *Fix an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$. Given any scalar $\gamma \in [0,1]$ and any two vectors $x, y \in \mathbb{R}^d$, at least one of the following holds:*

(i) $\|y - \gamma x\| - \|y - x\| \leq -\frac{1}{2}\|x - \gamma x\|$.

(ii) $\|y\| \geq \frac{1}{4}\|\gamma x\|$.

Proof As in the proof of Lemma 5 we assume (ii) does not hold and show that (i) does. WLOG, let $\|x\| = 1$. Let $\|\cdot\|_*$ denote the dual norm. Let $z_\tau := \nabla\|\tau x - y\| = \arg\max_{\|z\|_* \leq 1}\langle \tau x - y, z\rangle$ and note that $\langle z_\tau, \tau x - y\rangle = \|\tau x - y\|$. Then,

$$
\begin{aligned}
\frac{d}{d\tau}\|\tau x - y\| &= \left\langle \nabla\|\tau x - y\|, \frac{d}{d\tau}(\tau x - y)\right\rangle \\
&= \langle z_\tau, x\rangle \\
&= \frac{\langle z_\tau, \tau x - y\rangle + \langle z_\tau, y\rangle}{\tau} \\
&\geq \frac{\|\tau x - y\| - \|z_\tau\|_*\|y\|}{\tau} && \text{(definition of } z_\tau \text{ and Hölder)} \\
&\geq \frac{(\tau - \|y\|) - 1\cdot\|y\|}{\tau} = 1 - \frac{2\|y\|}{\tau}. && \text{(triangle inequality)}
\end{aligned}
$$

Given the bound $\frac{d}{d\tau}\|\tau x - y\| \geq 1 - \frac{2\|y\|}{\tau}$ we can say:

$$
\|y - \gamma x\| - \|y - x\| = -\int_\gamma^1 \frac{d}{d\tau}\left(\|\tau x - y\|\right) d\tau \leq -\int_\gamma^1 \left(1 - \frac{2\|y\|}{\tau}\right) d\tau. \tag{21}
$$

Since by assumption condition (ii) does hold and $\|x\| = 1$, we know that $\|y\| < \frac{1}{4}\|\gamma x\| = \frac{1}{4}\gamma$. Hence $\frac{2\|y\|}{\tau} < \frac{\gamma/2}{\tau} \leq 1/2$ for $\tau \geq \gamma$. The integrand in (21) is therefore at least half, and hence the result is at most $-\frac{1}{2}(1 - \gamma) = -\frac{1}{2}\|x - \gamma x\|$. Hence the proof. ∎