Lecture 12: Oracle Complexity Lower Bounds

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Convex Optimization

Prof. Zhiyuan Li

Spring 2026

What Is an Oracle Complexity Lower Bound?

The general form. Fix a function class $\mathcal{F}$ and an oracle. A *lower bound* says: for every algorithm A, there exists a hard $f \in \mathcal{F}$ that fools A.

$$\inf_{A} \sup_{f \in \mathcal{F}} \left[f(x_{T+1}) - f^\star \right] \geq \Delta(T).$$

Why must it be a class? Without a class, the statement is vacuous:

- Given a single f, the algorithm "output $\arg\min f$" has zero error.
- Lower bounds are inherently statements about *worst-case complexity over a family*.

What the class fixes.

- The objects f (e.g. convex G-Lipschitz, convex L-smooth, ...).
- The *oracle*: which information the algorithm may query (function value, subgradient, gradient, ...).
- The feasible region / radius (e.g. a Euclidean ball of radius R).

Why Lower Bounds? Certificates of Optimality

Course thread: certificates.

- Convex duality: a dual-feasible point certifies primal near-optimality.
- KKT: multipliers certify that a primal point is optimal.
- FW gap: $g_{\mathrm{FW}}(x)$ certifies $f(x) - f^\star \le g_{\mathrm{FW}}(x)$.

Lower bounds = certificates one level up. A lower bound $\Delta(T)$ certifies that *no* first-order algorithm can beat rate $\Delta(T)$ on $\mathcal{F}$.

Pairing it with a matching upper bound is exactly weak/strong duality at the algorithm level: a tight characterization of the class.

Without a lower bound. An $O(1/T)$ algorithm could be either optimal or off by a factor of T. We cannot tell whether to look for a better algorithm or accept the rate.

Today.

- Nonsmooth G-Lip., radius R: $\Omega(GR/\sqrt{T})$, tight against subgradient / mirror descent.
- Smooth L, radius R: $\Omega(LR^2/T^2)$; the algorithm matching this is Lecture 13.

First-Order Oracle and Method

Definition 12.1 (Subgradient oracle). For convex $f : \mathbb{R}^d \to \mathbb{R}$, a subgradient oracle returns on input x the pair

$$\mathsf{O}_f(x) = (f(x), g), \qquad g \in \partial f(x).$$

Definition 12.2 (Deterministic first-order method). A rule A mapping previous oracle replies to the next query:

$$x_1 = A(\varnothing), \qquad x_{t+1} = A\big((f(x_1), g_1), \ldots, (f(x_t), g_t)\big).$$

A run of length T uses T oracle calls and outputs x_{T+1}.

Note. "Deterministic" is essential: otherwise the algorithm could randomize over the hidden directions.

Zero-Initialized Linear-Span Method

Definition 12.3 (Zero-initialized linear-span method). A deterministic method with $x_1 = 0$ and

$$x_{t+1} \in \operatorname{span}\{g_1, \dots, g_t\}, \qquad t = 1, 2, \dots$$

Examples. Gradient descent from 0, conjugate gradient, mirror descent (Euclidean) from 0.

Especially natural under ℓ_2 constraints. If K is ℓ_2-symmetric (e.g. a Euclidean ball) and x_{t+1} has any component orthogonal to $\operatorname{span}\{g_1, \dots, g_t\}$:

- that component is invisible to the first-order oracle (no g_s certifies it);
- under $\|\cdot\|_2$ it only inflates $\|x_{t+1}\|_2$.

Projecting it away strictly improves the iterate. **Linear-span is the geometrically efficient choice in ℓ_2.** Under non-Euclidean norms this argument fails — which is why the lifting reduction below has to do real work in general.

From Method to Transcript

Definition 12.4 (Span-respecting transcript). Fix convex $h : \mathbb{R}^m \to \mathbb{R}$ and a subgradient oracle O_h. A transcript

$$(y_1, g_1), \ldots, (y_T, g_T), y_{T+1}$$

is *zero-initialized span-respecting* for (h, O_h) if

$$y_1 = 0, \quad \mathsf{O}_h(y_t) = (h(y_t), g_t), \quad y_{t+1} \in \operatorname{span}\{g_1, \ldots, g_t\}.$$

(The empty span is $\{0\}$.)

Why transcript, not method?

- Every zero-initialized linear-span method *produces* such a transcript.
- More general methods can be *forced into* one by hidden-subspace lifting.

Quantifier Order: $\exists$ vs $\forall$

Linear-span lower bound (today's first proof). Single hard instance defeats every transcript:

$$\exists\,(h, \mathrm{O}_h) \quad \forall\,\tau \text{ span-respecting}, \qquad h(y_{T+1}) - \min_{\|y\|_2 \le R} h(y) \ge \Delta.$$

Black-box lower bound (after the reduction). Hard instance depends on the method:

$$\forall\, A \quad \exists\,(f_U, \mathrm{O}_{f_U}), \qquad f_U(x_{T+1}) - \min_{\|x\|_2 \le R} f_U(x) \ge \Delta.$$

The bridge. Hidden-subspace *lifting*: a fixed low-dimensional hard instance, embedded in a method-dependent isometry $U : \mathbb{R}^m \to \mathbb{R}^D$, forces the projected ambient transcript to be span-respecting in $\mathbb{R}^m$.

Hard Instance: The Max-Coordinate Function

Lemma 12.1 (Max-coordinate is Lipschitz). For $G > 0$,

$$f(x) := G \max_{1 \le i \le d} x_i, \qquad x \in \mathbb{R}^d,$$

is convex and G-Lipschitz w.r.t. $\|\cdot\|_2$. For every $j \in \operatorname{argmax}_i x_i$, $Ge_j \in \partial f(x)$.

Why this instance?

- f is the pointwise max of linear functions $\Rightarrow$ convex.
- Subgradient at x is a coordinate basis vector Ge_j.
- One oracle call exposes *at most one new coordinate direction*.

Oracle choice (deterministic tie-breaking).

$$\mathrm{O}_h(y) := \big(h(y), Ge_{j(y)}\big), \qquad j(y) := \min \operatorname*{argmax}_{1 \le i \le m} y_i.$$

Nonsmooth Lower Bound (Linear-Span Methods)

Theorem 12.2 (Fixed max-coordinate lower bound). Let $T \in \mathbb{N}$, $G, R > 0$, and set $m := T + 1$. Define

$$h(y) := G \max_{1 \le i \le m} y_i, \qquad \mathrm{O}_h(y) := \big(h(y), G e_{j(y)}\big), \quad j(y) := \min \operatorname{argmax}_i y_i.$$

Then every length-T zero-initialized span-respecting transcript for (h, O_h) satisfies

$$\boxed{h(y_{T+1}) - \min_{\|y\|_2 \le R} h(y) \;\ge\; \frac{GR}{\sqrt{T+1}}.}$$

Consequence. Any zero-initialized linear-span method needs $T + 1 \ge G^2 R^2 / \varepsilon^2$ oracle calls to drive error below ε.

Proof of Theorem 12.2 (1/2): Missing Coordinate

Step 1. The oracle returns $g_t = Ge_{j_t}$, where $j_t := j(y_t) = \min \operatorname{argmax}_i (y_t)_i$. So every subgradient is a scalar multiple of a standard basis vector.

Step 2. Span has at most T directions. Span-respecting:

$$y_{T+1} \in \operatorname{span}\{g_1, \ldots, g_T\} = \operatorname{span}\{e_{j_1}, \ldots, e_{j_T}\}.$$

At most T coordinate directions are used, but $m = T + 1$. **Some index k is missed.**

Step 3. Max coordinate of y_{T+1} is nonnegative. Since $(y_{T+1})_k = 0$,

$$h(y_{T+1}) = G \max_{1 \le i \le m} (y_{T+1})_i \ge G(y_{T+1})_k = 0.$$

Role of convexity / G-Lipschitzness. These are class membership constraints: the hard instance *must* be convex and G-Lipschitz to live in the function class for which we're proving the lower bound. They are not tools used by the proof.

Proof of Theorem 12.2 (2/2): The Comparator

Step 4. Pick a comparator with small max-coordinate. Let

$$y^\star := -\frac{R}{\sqrt{m}}\mathbf{1} \in \mathbb{R}^m, \qquad \mathbf{1} = (1, \ldots, 1)^\top.$$

Then $\|y^\star\|_2 = \sqrt{m \cdot R^2/m} = R$, so $y^\star$ is feasible. Its max coordinate is

$$h(y^\star) = G \cdot \left(-\frac{R}{\sqrt{m}}\right) = -\frac{GR}{\sqrt{m}}.$$

Step 5. Combine. Using $h(y_{T+1}) \geq 0$:

$$h(y_{T+1}) - \min_{\|y\|_2 \leq R} h(y) \;\geq\; h(y_{T+1}) - h(y^\star) \;\geq\; 0 + \frac{GR}{\sqrt{m}} = \frac{GR}{\sqrt{T+1}}.$$

□

Heart of the argument. The hard instance is *symmetric across coordinates*, so the optimum spreads its mass over all coordinates ($-(R/\sqrt{m})\mathbf{1}$); but any zero-initialized span-respecting iterate is supported on at most T coordinates and cannot reach this symmetric optimum.

Why a Reduction Is Needed

What Theorem 12.2 proves. A bound for span-restricted transcripts.

What it does not prove. An arbitrary deterministic method might:

- start at $x_1 \neq 0$;
- make queries with components orthogonal to $\mathrm{span}\{g_1, \ldots, g_t\}$;
- exploit knowledge of the function class to predict the hidden coordinate.

The reduction (today's hardest proof). *Hidden-subspace lifting* transforms the linear-span lower bound into a fully unrestricted lower bound:

$$\begin{array}{c} \text{fixed low-dim instance } (h, \mathsf{O}_h) \\ + \text{ method-dependent isometry } U : \mathbb{R}^m \to \mathbb{R}^D \\ \Longrightarrow \text{ lifted instance } f_U(x) = h(U^\top x) \text{ in } \mathbb{R}^D \end{array}$$

Slogan. The method may probe arbitrary directions in $\mathbb{R}^D$, but f_U ignores everything outside the hidden m-dim subspace.

The Lifting Theorem

Theorem 12.3 (Hidden-subspace lifting). Let $T, m \in \mathbb{N}$, $D \geq m + T + 1$, $R, \Delta > 0$. Suppose (h, O_h) has the span-respecting bound $h(y_{T+1}) - \min_{\|y\|_2 \leq R} h(y) \geq \Delta$ for every length-T transcript. Then for every deterministic method A in $\mathbb{R}^D$, there is a linear isometry $U : \mathbb{R}^m \to \mathbb{R}^D$ such that

$$f_U(x) := h(U^\top x), \qquad \mathsf{O}_{f_U}(x) := \big(h(U^\top x),\ Ug(U^\top x)\big)$$

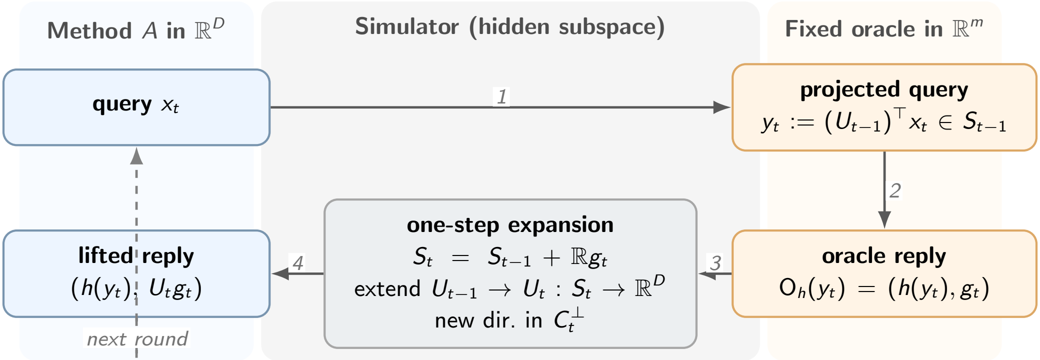
forces $f_U(x_{T+1}) - \min_{\|x\|_2 \leq R} f_U(x) \geq \Delta$.

Quantifier reversal. "$\exists (h, \mathsf{O}_h)\ \forall \tau$" becomes "$\forall A\ \exists (f_U, \mathsf{O}_{f_U})$".

Inheritance separated. Convexity, G-Lipschitzness, L-smoothness, the ball minimum, and the centered minimizer $x^\star = Uy^\star$ are routine facts about composing with a linear isometry — they are bundled into a separate lemma (**Lemma 12.4**) so the lifting theorem is just the existence of U plus the main lower-bound inequality.

Lifting Proof: The Simulator Picture

Build U adaptively while simulating A with the lifted oracle. *Round t:*



Key invariant. $y_t \in S_{t-1}$ holds *by definition* of $y_t = (U_{t-1})^\top x_t$. The work is: choosing U_t on the at-most-one-dim new piece $S_t \cap S_{t-1}^\perp$ so that $(U_q)^\top x_q = y_q$ stays consistent for $q \le t$.

Lifting Proof (1/2): Online Construction of U_t

Round t**.** Have $S_{t-1} = \mathrm{span}\{g_1, \ldots, g_{t-1}\}$ and partial isometry $U_{t-1} : S_{t-1} \to \mathbb{R}^D$. Method A outputs query x_t.

Project using the existing isometry.

$$y_t := (U_{t-1})^\top x_t \in S_{t-1}.$$

This is automatic — $(U_{t-1})^\top : \mathbb{R}^D \to S_{t-1}$ already maps into S_{t-1}.

Oracle reply. $\mathrm{O}_h(y_t) = (h(y_t), g_t)$. Set $S_t := S_{t-1} + \mathbb{R} g_t$ (*at most one new direction*).

Extend $U_{t-1} \to U_t$ **on** S_t**.** Let $C_t := \mathrm{span}\big(\{x_1, \ldots, x_t\} \cup U_{t-1}(S_{t-1})\big)$. Dimension count:

$$\dim C_t^\perp \geq D - (t + r_{t-1}) \geq \dim(S_t \cap S_{t-1}^\perp) \quad (\text{uses } D \geq m + T + 1).$$

Choose $U_t : S_t \to \mathbb{R}^D$ with $U_t|_{S_{t-1}} = U_{t-1}$ and $U_t(S_t \cap S_{t-1}^\perp) \subseteq C_t^\perp$.

Past-query consistency. The new direction is orthogonal to all x_q ($q \leq t$), so $(U_t)^\top x_q = (U_{t-1})^\top x_q$. By induction $(U_t)^\top x_q = y_q$ for $q \leq t$.

Lifting Proof (2/2): Completion + Lemma 12.4

After round T. Method outputs x_{T+1}. Same dimension count $\Rightarrow$ extend U_T to a full isometry

$$U : \mathbb{R}^m \to \mathbb{R}^D, \qquad U = U_T \text{ on } S_T, \quad U(S_T^\perp) \perp \{x_1, \ldots, x_{T+1}\}.$$

Then $U^\top x_t = y_t$ for $t \le T$ and $U^\top x_{T+1} \in S_T$. The projected transcript $(y_1, g_1), \ldots, (y_T, g_T), y_{T+1}$ is span-respecting for (h, O_h).

Lemma 12.4 (Lifted composition preserves regularity, minimizers, gaps). For any linear isometry $U : \mathbb{R}^m \to \mathbb{R}^D$, $f_U := h \circ U^\top$ satisfies:

- convex; O_{f_U} is a valid subgradient oracle;
- $\inf_{\|x\|_2 \le R} f_U(x) = \inf_{\|y\|_2 \le R} h(y)$;
- h is G-Lip. / L-smooth $\Rightarrow$ so is f_U;
- if $y^\star \in \operatorname{argmin} h$, then $x^\star := Uy^\star \in \operatorname{argmin} f_U$, $\|x^\star\|_2 = \|y^\star\|_2$.

Conclude. The span-respecting bound on h + Lemma 12.4 $\Rightarrow$
$f_U(x_{T+1}) - \min_{\|x\|_2 \le R} f_U(x) = h(y_{T+1}) - \min_{\|y\|_2 \le R} h(y) \ge \Delta$. $\square$

Unrestricted Nonsmooth Lower Bound

Theorem 12.5 (Black-box nonsmooth). Let $T \in \mathbb{N}$, $G, R > 0$, $D \geq 2T + 2$. For every deterministic method A in $\mathbb{R}^D$, there exist a G-Lipschitz convex $f : \mathbb{R}^D \to \mathbb{R}$ and a subgradient oracle for f such that

$$f(x_{T+1}) - \min_{\|x\|_2 \leq R} f(x) \geq \frac{GR}{\sqrt{T+1}}.$$

Proof. Apply Theorem 12.3 (lifting) with $m = T + 1$ to the fixed instance from Theorem 12.2. The dimension assumption $D \geq 2T + 2 = m + T + 1$ is exactly what the lifting theorem needs. □

Matching upper bound. Subgradient method / mirror descent achieves $O(GR/\sqrt{T})$ (Lecture 9). **Both rates tight up to constants.**

Smooth Lower Bound: Quadratic Chain

Hard instance. Set $d := T + 1$. For $\alpha = R/\sqrt{d}$,

$$h(y) := \frac{L}{8}\, y^\top B y - \frac{L\alpha}{4}\, e_1^\top y, \quad y^\top B y = y_1^2 + \sum_{i=1}^{d-1} (y_i - y_{i+1})^2.$$

Key properties.

- $B \succeq 0$, $\|B\|_{\mathrm{op}} \le 4 \Rightarrow h$ is convex, L-smooth.
- $B\mathbf{1} = e_1$, so $y^\star := \alpha\mathbf{1}$ minimizes h and $\|y^\star\|_2 = R$.
- B is a tridiagonal "chain": $B_{ij} \neq 0 \Leftrightarrow |i - j| \le 1$.

Theorem 12.6 (Smooth zero-init span-respecting bound). Every length-T zero-init span-respecting transcript for h with the exact gradient oracle satisfies

$$\boxed{h(y_{T+1}) - h(y^\star) \;\ge\; \frac{LR^2}{8(T+1)^2}.}$$

Proof of Theorem 12.6 (1/3): h Is the Right Object

Step 1. Operator norm of B. Expanding the quadratic form,

$$y^\top By \le y_1^2 + 2\sum_{i=1}^{d-1}(y_i^2 + y_{i+1}^2) \le 4\|y\|_2^2,$$

so $\|B\|_{\mathrm{op}} \le 4$. Hence $\nabla^2 h = (L/4)B$ has $\|\nabla^2 h\|_{\mathrm{op}} \le L$.

Step 2. Convexity and smoothness. $\nabla^2 h \succeq 0$ and $\|\nabla^2 h\|_{\mathrm{op}} \le L$:

$$h \text{ is convex and } L\text{-smooth w.r.t. } \|\cdot\|_2.$$

Step 3. Minimizer. $\nabla h(y) = (L/4)(By - \alpha e_1)$, so the minimizer solves $By = \alpha e_1$. On the constant vector $y^\star = \alpha\mathbf{1}$, all difference terms vanish:

$$B\mathbf{1} = \big(2\cdot 1 - 1 + 1,\ 1-1, \ldots, 1-1\big)^\top \cdot (?) \implies B\mathbf{1} = e_1.$$

(boundary term contributes 1 to coord. 1; chain differences are zero.) Hence $y^\star = \alpha\mathbf{1} \in \operatorname{argmin} h$ with $\|y^\star\|_2 = R$.

Proof of Theorem 12.6 (2/3): Information Spreads One Coordinate at a Time

Sparsity propagation. If $\mathrm{supp}(x) \subseteq \{1, \ldots, k\}$, the chain structure of B gives

$$\mathrm{supp}(Bx) \subseteq \{1, \ldots, k+1\}.$$

Since $\nabla h(x) = (L/4)(Bx - \alpha e_1)$, the same is true for the gradient.

Inductive support bound. Starting from $y_1 = 0$, span-respecting gives, by induction on t:

$$\mathrm{supp}(y_t) \subseteq \{1, \ldots, t-1\} \qquad \text{for } t = 1, \ldots, T+1.$$

In particular.

$$(y_{T+1})_d = 0, \qquad d = T+1.$$

The last coordinate of y_{T+1} is forced to be **exactly zero**.

Information-hiding picture. Each gradient query advances the "support frontier" by exactly one coordinate. After T queries, the iterate is supported on at most T coordinates — one short of d.

Proof of Theorem 12.6 (3/3): Discrete Poincaré on a Chain

Goal. Lower-bound the suboptimality $h(y_{T+1}) - h(y^\star)$.

Reduce gap → quadratic form. h quadratic, $\nabla h(y^\star) = 0$:

$$h(y_{T+1}) - h(y^\star) = \tfrac{L}{8} z^\top B z, \qquad z := y_{T+1} - y^\star.$$

It suffices to lower-bound $z^\top Bz$.

Boundary forced by Step 2. $(y_{T+1})_d = 0$ and $y^\star = \alpha\mathbf{1}$, so $z_d = -\alpha$. Set $z_0 := 0$ for convenience. Now z is a "path" from $z_0 = 0$ to $z_d = -\alpha$.

$z^\top Bz$ **= discrete Dirichlet energy of this path.**

$$z^\top Bz \;=\; \underbrace{z_1^2}_{=(z_1 - z_0)^2} + \sum_{i=1}^{d-1} (z_i - z_{i+1})^2 \;=\; \sum_{i=0}^{d-1} (z_{i+1} - z_i)^2.$$

A weighted sum of squared step sizes along the chain $0 \to 1 \to \ldots \to d$.

Cauchy–Schwarz: spreading α over d steps costs at least α^2/d.

$$\alpha^2 = (z_0 - z_d)^2 = \left(\sum_{i=0}^{d-1} (z_i - z_{i+1})\right)^2 \le d \sum_{i=0}^{d-1} (z_{i+1} - z_i)^2 = d\, z^\top Bz.$$

Unrestricted Smooth Lower Bound

Theorem 12.7 (Black-box smooth). Let $T \in \mathbb{N}$, $L, R > 0$, $D \geq 2T + 2$. For every deterministic method A in $\mathbb{R}^D$, there exist a convex L-smooth quadratic $f : \mathbb{R}^D \to \mathbb{R}$, its exact gradient oracle, and $x^\star \in \operatorname{argmin} f$ with $\|x^\star\|_2 \leq R$ such that

$$f(x_{T+1}) - f(x^\star) \geq \frac{LR^2}{8(T+1)^2}.$$

Proof. Apply lifting (Theorem 12.3) with $\Delta = LR^2/8(T+1)^2$ and $m = d = T + 1$. Lemma 12.4 then gives a smooth f_U with minimizer $x^\star = Uy^\star$ satisfying $\|x^\star\|_2 = \|y^\star\|_2 = R$. □

Where we are vs. this bound. Gradient descent gives only $O(LR^2/T)$ — a factor of T off. The matching $O(LR^2/T^2)$ algorithm is the goal of Lecture 13.

Rate Summary (I): What We've Proved So Far (Euclidean)

Setting (Euclidean)	**Lower bound**	**Upper bound (proved in this course)**
Nonsmooth G-Lip., radius R	$\Omega(GR/\sqrt{T})$ [Thm 12.5]	✓ $O(GR/\sqrt{T})$ (subgradient / mirror descent)
Nonsmooth μ-str. cvx, G-Lip.	$\Omega(G^2/(\mu T))$	✓ $O(G^2/(\mu T))$ (weighted mirror descent)
Smooth L, radius R	$\Omega(LR^2/T^2)$ [Thm 12.7]	$\times$ $O(LR^2/T)$ (GD) — gap of T
Smooth L, μ-str. cvx, $\kappa = L/\mu$	$\Omega(\sqrt{\kappa}\log(1/\varepsilon))$	$\times$ $O(\kappa \log(1/\varepsilon))$ (GD) — gap of $\sqrt{\kappa}$
Stoch. nonsmooth, 2nd-mom G, radius R	$\Omega(GR/\sqrt{T})$	✓ $O(GR/\sqrt{T})$ (SMD)
Stoch. smooth L, var. σ^2, radius R	$\Omega(LR^2/T^2 + \sigma R/\sqrt{T})$	$\times$ $O(LR^2/T + \sigma R/\sqrt{T})$ (SMD) — gap on smooth term

Green ✓ matches the lower bound; Red $\times$ has an open gap with the algorithms in this course so far.

Rate Summary (II): Acceleration Closes the Smooth Gap

Setting (Euclidean)	**Lower bound**	**Upper bound (with acceleration)**
Nonsmooth G-Lip., radius R	$\Omega(GR/\sqrt{T})$	✓ $O(GR/\sqrt{T})$ (subgradient / mirror descent)
Nonsmooth μ-str. cvx, G-Lip.	$\Omega(G^2/(\mu T))$	✓ $O(G^2/(\mu T))$ (weighted mirror descent)
Smooth L, radius R	$\Omega(LR^2/T^2)$	✓ $O(LR^2/T^2)$ (**accelerated GD — Lecture 13**)
Smooth L, μ-str. cvx, $\kappa = L/\mu$	$\Omega(\sqrt{\kappa}\log(1/\varepsilon))$	✓ $O(\sqrt{\kappa}\log(1/\varepsilon))$ (**accelerated — Lecture 13**)
Stoch. nonsmooth, 2nd-mom G, radius R	$\Omega(GR/\sqrt{T})$	✓ $O(GR/\sqrt{T})$ (SMD)
Stoch. smooth L, var. σ^2, radius R	$\Omega(LR^2/T^2 + \sigma R/\sqrt{T})$	✓ $O(LR^2/T^2 + \sigma R/\sqrt{T})$ (AC-SA, Lan 2012)

In Lecture 13 we will construct *accelerated gradient methods* that attain the smooth Euclidean rates above.

Rate Summary (III): Beyond Euclidean — Acceleration Is Not Free

Setting	Lower bound	Upper bound
Smooth on ℓ_p-ball, $2 \leq p < \infty$, (s, L)-Hölder smooth in $\|\cdot\|_p$, $T \leq n$	$\Omega\left(\frac{LR^s}{(\min\{p, \log T\})^{s-1}\ T^{s+s/p-1}}\right)$	✓ $O\left(\frac{LR^s}{T^{s+s/p-1}}\right)$ (non-Euclidean accel.)
Smooth on ℓ_∞-ball, L-smooth in $\|\cdot\|_\infty$	$\Omega(LR^2/(T \log T))$	✓ $O(LR^2/T)$ (Frank–Wolfe) — off by $\log T$ only

Acceleration does not always happen.

- ℓ_p with $2 \leq p < \infty$: non-Euclidean acceleration matches the T-rate, but the constant degrades as $(\min\{p, \log T\})^{s-1}$. **Larger $p \Rightarrow$ less room for acceleration.**
- ℓ_∞ ($p = \infty$ limit): geometry kills acceleration entirely. Best known is Frank–Wolfe at $O(LR^2/T)$, off by $\log T$ from the lower bound.

Summary

The information-hiding paradigm.

- One oracle call exposes one new direction (max-coordinate / chain).
- After T calls, one direction or coordinate remains unseen.
- A symmetric / boundary-supported optimum exploits that gap.

Two regimes, one paradigm.

- Nonsmooth: $\Omega(GR/\sqrt{T})$ via the max-coordinate function.
- Smooth: $\Omega(LR^2/T^2)$ via the quadratic chain.

From span-restricted to black-box. Hidden-subspace lifting:

$$\exists (h, \mathsf{O}_h)\, \forall \tau \implies \forall A\, \exists (f_U, \mathsf{O}_{f_U}).$$

Adaptive partial isometry U_t, finalized after the run.

Next. Lecture 13 closes the smooth Euclidean gap: an algorithm matching $\Omega(LR^2/T^2)$.

Bibliographic Notes

- **Nemirovski & Yudin (1983).** Foundational text on oracle complexity; first-order lower bounds in the linear-span and black-box settings.
- **Nesterov (Introductory Lectures, 2004).** Theorems 2.1.7 (smooth convex $\Omega(LR^2/T^2)$) and 2.1.13 (smooth strongly convex $\Omega(\sqrt{\kappa}\log(1/\varepsilon))$) used as references in the rate table.
- **Agarwal, Bartlett, Ravikumar & Wainwright (2012).** Information-theoretic stochastic oracle lower bounds.
- **Lan (Math. Program. 2012).** AC-SA achieves the optimal stochastic smooth rate $LR^2/T^2 + \sigma R/\sqrt{T}$.
- **Braun, Guzmán & Pokutta (IEEE TIT 2017).** Distributional oracle complexity for randomized nonsmooth black-box methods.
- **Diakonikolas & Guzmán (COLT 2020).** Parallel and randomized local-oracle lower bounds.
- **Guzmán & Nemirovski (J. Complexity 2015).** Smooth convex lower bounds on ℓ_p and ℓ_∞ balls.