Summary Optimization II

Contents

1 Introduction

Theorem: Extreme value/Weierstrass

Let $f : \Omega \to \mathbb{R}$ be continuous and $\Omega \subset \mathbb{R}^n$ compact. Then f attains its min and max on Ω.

Definition: Coercive

$f : \mathbb{R}^n \to \mathbb{R}$ is called coercive if for every sequence $(x^{(k)})$ for which $\|x^{(k)}\| \underset{k\to\infty}{\longrightarrow} \infty$ then $f(x^{(k)}) \to \infty$

Theorem: Existence min

Let $f : \mathbb{R}^n \to \mathbb{R}$ be coercive and continuous, then f has a global min.

Theorem: Midpoint convexity

A closed midpoint convex set is also convex.

Operations that conserve convexity

- ⋆ Intersection (may be infinite)
- ⋆ Projection onto coordinate
- ⋆ Cartesian product
- ⋆ Minkowski sum: $S_1 + S_2 := \{x + y \mid x \in S_1,\ y \in S_2\}$
- ⋆ Affine transformations
- ⋆ If $f : \mathbb{R}^d \to \mathbb{R}$ is convex, then $f(Ax + b)$ is convex.
- ⋆ If $f_i : \Omega \to \mathbb{R}$ are convex and $\omega_i \geq 0$, then this is convex:

$$f(x) := \sum_{i=1}^{m} \omega_i f_i(x) \qquad (\forall m \in \mathbb{N} \cup \{+\infty\})$$

- ⋆ If $f_i : \Omega \to \mathbb{R}$ are convex, then $f(x) := \sup_{i=1,\ldots,m} f_i(x)$ is convex $(m \in \mathbb{N} \cup \infty)$
- ⋆ If f is convex and non-negative, then $(f(x))^\alpha$ is convex $\forall \alpha \geq 1$
- ⋆ If $f(x, y)$ is convex on $\mathbb{R}^{d_1+d_2}$ and $Y \subset \mathbb{R}^{d_2}$ is convex, then

$$g(x) := \inf_{y\in Y} f(x, y) \quad \text{is convex}$$

Thm about convex hull

Let $S \subset \mathbb{R}^d$. Ten every points in conv S can be written as a convex combination of at most $d + 1$ points in S.

2 Convex Programming

Thm: equivalence of problem

Let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a linear functioin and $\Omega \subset \mathbb{R}^n$ be compact.
Then $\min \ell(x)$ s.t. $x \in \Omega$ is equivalent to: $\min \ell(x)$ s.t. $x \in \text{conv}(\Omega)$.

Definition: α-strong convexity

f is called α-strong convex if $g(x) := f(x) - \frac{\alpha}{2}\|x\|_2^2$ is convex ($\alpha > 0$).

Note: Strong convex $\implies$ strict convex $\implies$ convex
Some reformulation:

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \alpha \|x - y\|^2$$
$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\alpha}{2}\|x - y\|^2$$

Thm: Grad on convex functions

Let $f : \mathbb{R}^n \to R$ be given. If f is differentiable, then the following are equivalent:

1. f is convex
2. $f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in \mathbb{R}^n$
3. $\nabla^2 f(x) \geq 0 \quad \forall x \in \mathbb{R}^n$

Note: the function $f(X) = \log\det(X)$ is concave for $X \in \mathbb{S}_+^n$

Optimality conditioins

Consider $\min\limits_{x \in \mathbb{R}^d} f(x)$ s.t. $x \in \Omega$ where Ω, f are conved and f is differentiable.

$$x^* \text{ is optimal } \iff x^* \in \Omega \text{ and } \nabla f(x^*)^T (y - x^*) \geq 0 \quad \forall y \in \Omega$$

3 Standard Problems

Linear Programming (LP)

An LP is the problem of optimizing a linear function over a polyhedron:

$$\min_{x\in\mathbb{R}^n} c^T x \quad c\in\mathbb{R}^n \quad \text{s.t.} \quad a_i^T x \le b_i \quad i=1\dots m \quad \text{equiv:} \quad Ax\le b$$

Note: Standard form of LP:

$$\min_{x\in\mathbb{R}^n} c^T x \quad \text{s.t.} \quad Bx=d,\ x\ge 0$$

Also, the max/min flow problem is a linear problem.

Quadratic programming (QP)

A QP is optimizing a quadratic function over a polyhedron:

$$\min_{x\in\mathbb{R}^n} x^T Qx + q^t x\ (+c) \quad \text{s.t.} \quad Ax\le b \quad Q\in\mathbb{S}^n$$

The QP is convex iff $Q\succeq 0$ otherwise it's non-convex and quite difficult to solve.

Quadraticly constrained QP (QCQP)

$$minx^t Q_0 x + q_0^T x + c_0 \quad \text{s.t.} \quad x^T Q_i x + q_i^T x + c_i \le 0 \quad i=1\dots m$$

This is convex if $Q_0,\dots,Q_m\succeq 0$

Second-order cone programming (SOCP)

$$\min_x q^T x \quad \text{s.t.} \quad \|A_i x + b_i\|_2 \le c_i^T x + d_i \quad i=1\dots m,\ A_i\in\mathbb{R}^{\ell_i\times n}$$

Note: QCQP $\subsetneq$ SOCP

Geometric Programming (GP)

Monomial function: $f(\vec{x}) = c\cdot x_1^{a_1}\cdots x_n^{a_n}$ with $c>0,\ a_i\in\mathbb{R},\ x_i>0$.

Posinomial function: $g(\vec{x}) = \sum_{i=1}^{m} f_i(\vec{x})$ where f_i are monomials.

The posinomials are closed under addition and multiplication, the monomials are closed under multiplication and division.

Def GP:

$$\min_{x_i>0} f_0(\vec{x}) \quad \text{s.t.} \quad f_i(\vec{x})\le 1,\ h_j(\vec{x})=1$$

where f_0 is a posy., f_i are posy. and h_j are mono.

This is generally not convex but with change of variables and passing by the fact that the log-sum-exp function is convex:

$$f(x) = \log(\sum_i \exp(x_i))$$

We can show that GP is equivalent to a convex programming problem.

Semidefinite programming (SDP)

$$\min_{X\in\mathbb{S}} \mathrm{Tr}(C\cdot X) \quad \text{s.t.} \quad \mathrm{Tr}(A_i X) = b_i,\ X\succeq 0, \quad C, A_i\in\mathbb{S}$$

This formulation is a generalisation of LP

Support vector machine (SVM)

The role of this programming is to "classify" data into categories. For the simplest case, a linear sepparation we can set the problem with the following formulation:

$$\max_{a,b,t} t \quad \text{s.t.} \quad y_i(a^T x_i + b) \geq t \quad \forall i,\ \|a\|_2 = 1$$

4 Duality

4.1 Weak duality

Lagrangian

Take a programming problem:

$$\min_{x\in\mathbb{R}^n} f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0 \quad i = 1\dots m \quad \text{and} \quad h_j(x) = 0 \quad j = 1\dots p \tag{P}$$

The Lagragian $\mathcal{L}:\ D \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is defined:

$$(\vec{x}, \vec{\lambda}, \vec{\nu}) \mapsto f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x})$$

Note that $D \subset \mathbb{R}^n$: $D = (\cap \text{ dom } f_i) \cap (\cap \text{ dom } h_j)$ The Lagrangian dual function of (P) is: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$

$$(\vec{\lambda}, \vec{\nu}) \mapsto \inf_{x\in D} \mathcal{L}(x, \lambda, \nu)$$

Note that g is a concave function !!

Thm: Lower bound

The dual function g gives a lower bound to the optimal value of (P):

$$\forall \lambda \geq 0,\ \nu \in \mathbb{R}^p \implies g(\lambda, \nu) \leq p_*$$

Remarque: In general the max of $g(\lambda, \nu)$ gives the best lower bound for the primal problem.

Some duals

- Note that the dual of LP in standard form can be computed to be an LP in inequality form and the reverse also hold.
- The dual of an SDP is an SDP in LMI form

4.2 Strong duality

We know that for weak duality $d_* \leq p_*$, strong duality is when the equality hold, i.e.: $d_* = p_*$. This is true for most convex problem and rarely for non convex problem.

Slater's condition

Let (P) be the problem: $\min_x f_0(x)$ s.t. $f_i(x) \leq 0$, $i = 1\dots m$ where f_i are convex and $Ax = b$. Denote $\mathcal{A} := \{$ index i, such that f_i is affine $\}$

The program (P) satisfies Slater'scondition if it is strictly feasible, i.e:

$$\exists \tilde{x} \in \text{ rel int } D \quad \text{s.t.} \quad f_i(\tilde{x}) < 0,\ A\tilde{x} = b$$

It satisfies the weak Slater's condition is:

$$\exists \tilde{x} \in \text{ rel int } D \quad \text{s.t.} \quad f_i(\tilde{x}) \begin{cases} \leq 0 & i \in \mathcal{A} \\ < 0 & \text{Otherwise} \end{cases},\ A\tilde{x} = b$$

Thm: Weak Slater's

If (P) is convex and satisfis the weak Slater's condition, then strong duality holds.

Thm: Separation Hyperplane

Let $C, D \subset \mathbb{R}^n$ be two disjoint, convex, non empty sets. Then $\exists a \in \mathbb{R}^n,\ a \neq 0$ s.t.:

$$\inf_{x \in D} a^T x \geq \sup_{x \in C} a^t x$$

(In other words: there exists an hyperplane separating the two sets)

Thm: Convex strong duality

Let (P) be convex, bouded below, and satisfying slater's condition. Then strong duality holds.

4.3 Saddle points

Def: weak max-min property

$$\sup_{y \in Y} \inf_{x \in X} \phi(x, y) \leq \inf_{x \in X} \sup_{y \in Y} \phi(x, y)$$

Note: This holds for any function $\phi :\ X \times Y \to \mathbb{R}$

Def: Saddle point

A function $\phi :\ X \times Y \to \mathbb{R}$ has a saddle point if $\exists x^* \in X$ and $y^* \in Y$ such that:

$$\phi(x^*, y) \leq \phi(x^*, y^*) \leq \phi(x, y^*) \qquad \forall x \in X,\ y \in Y$$

Thm: Strong max-min property

If ϕ has a saddle point, then it will satisfy the strong max-min property:

$$\sup_{y \in Y} \inf_{x \in X} \phi(x, y) = \inf_{x \in X} \sup_{y \in Y} \phi(x, y)$$

Thm: Saddle/Strong duality

The set of saddle points of the Lagrangian of (P) is non empty iff strong duality holds for (P). And in that case, the optimal primal/dual solutions are the saddle points.

KKT condition for (P)

$$\exists x^*, \lambda^* \quad \text{s.t.} \quad x^* \in D, \lambda^* \geq 0 \quad \text{and} \quad \nabla_x \mathcal{L}(x^*, \lambda^*) = 0,\ f_i(x^*) \leq 0,\ \lambda_i f_i(x^*) = 0$$

5 Algorithms for unconstrained opt. prob.

5.1 On function smoothness

Def: β-smoothness

A function $f : \mathbb{R}^n \to \mathbb{R}$ is β-smooth if $\nabla f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with parameter β:

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\| \qquad \forall x, y \in \mathbb{R}^n$$

Note: If $f \in C^2$, then β-smoothness is equiv to $\nabla^2 f(x) \leq \beta \cdot Id$

Lemma on β-smoothness

$$|f(y) - f(x) - \nabla f(x)^T (y - x)| \leq \frac{\beta}{2} \|y - x\|^2 \qquad \forall x, y \in \mathbb{R}^n$$

Def: Type (α, β)

A function is of type (α, β) if it is α strongly convex and β-smooth.

Thm: On (α, β)

Let f be of type (α, β), then:

$$\frac{1}{2\beta} \|\nabla f(x)\|_2^2 \leq f(x) - p_* \leq \frac{1}{2\alpha} \|\nabla f(x)\|_2^2$$

Notes: Hence up to a factor $\frac{1}{2\beta}$ and $\frac{1}{2\alpha}$, the gradiant squared is comparable to error in function value.
$\longrightarrow$ If we have a point x with small gradiant, then the error in function value at this point is also small.

Thm: Error bound on x

Let f be of type (α, β), then:

$$\|x - x_*\|_2 \leq \frac{2}{\alpha} \|\nabla f(x)\|_2$$

5.2 Gradient method

Idea:

$$\vec{x_{k+1}} = \vec{x_k} + t_k \cdot \vec{d_k}, \; t_k \quad \text{the stepsize and} \quad \vec{d_k} \quad \text{the direction}$$

We want to find a good direction (and stepsize) suck that $f(x_{k+1}) < f(x_k)$. When we chose $d_k = \nabla f(x_k)$ as direction, we get steepest descent.

Thm: Conv speed SD

Let f be of type (α, β). Then SD with exact line search satisfies:

$$f(x_k) - p_* \leq \left(1 - \frac{\alpha}{\beta}\right)^k (f(x_0) - p_*)$$

Thm: Better conv. bound

Let f be of type (α, β). Then SD with constant stepsize $t = \dfrac{2}{\alpha + \beta}$ satisfies:

$$\|x_k - x_*\|_2 \leq \tilde{\gamma}^k \|x_0 - x_*\|_2 \qquad \tilde{\gamma} = 1 - \frac{2\alpha}{\alpha + \beta} = \frac{\kappa - 1}{\kappa + 1}, \quad \kappa = \frac{\beta}{\alpha}$$

Remarque: We can retrieve the bound in function value using β-smoothness:

$$\begin{aligned} f(x_k) - f(x_*) &\leq \cancelto{0}{\nabla f(x_*)^T (x_k - x_*)} + \frac{\beta}{2}\|x_k - x_*\|_2^2 \\ &\leq \frac{\beta}{2}\tilde{\gamma}^{2k}\|x_0 - x_*\|^2 \end{aligned}$$

5.3 Oracle Complexity

Idea of this section is to study generaly the quality of methods. For this purpose we define an oracle. Said simply, an oracle is a little machine that, given an imput x, will output certains information about our function f. Ex: 0th order oracle will only output $f(x)$, 1st order oracle will also output $\nabla f(x)$, 2th order adds $\nabla^2 f(x)$ etc. In this course we'll limit ourselves at 0th and 1st order miracles.

We define the cost of a methad as the number of times it queries the oracle. The method computes better points based on these oracle calls:

(1) Given x_0, obtain $f(x_0), \nabla f(x_0)$.
Compute x_1 based on $f(x_0), \nabla f(x_0)$: $x_1 = \phi_1(x_0, f(x_0), \nabla f(x_0))$

(2) Given x_1, obtain $f(x_1), \nabla f(x_1)$.
Compute x_2 based on the history: $x_2 = \phi_2(x_0, f(x_0), \nabla f(x_0), x_1, f(x_1), \nabla f(x_1))$

(k) compute $x_k = \phi_k(x_0, \ldots, x_{k-1}, f(x_0), \ldots, f(x_{k-1}), \nabla f(x_0), \ldots, \nabla f(x_{k-1}))$

Def: Minimax oracle complexity

For a function class $\mathcal{F}$, e.g. convex of type (α, β), we define the minimax oracle complexity as:

$$\mathrm{OC}_k(\mathcal{F}) := \inf_{\phi_1, \ldots, \phi_{k-1}} \sup_{f \in \mathcal{F}} \big(f(x_k) - p_*\big)$$

In a more verbose way, this study the best possible outcome of our algorithm $\phi_1, \ldots, \phi_{k-1}$ for the worst possible function in $\mathcal{F}$. We cannot solve this directly but we can boud it to get an estimate:

$$(\star) \leq \sup_f \inf_{\phi_i} (f(x_k) - p_*) \leq \inf_{\phi_i} \sup_f f(x_k) - p_* \leq (\star\star)$$

($\star$) The error of any method applied to an existing function $\tilde{f} \in \mathcal{F}$.

($\star\star$) Error of an existing method $\tilde{\phi}$ applied to any $f \in \mathcal{F}$.

Since we're interested in the study of (α, β) functions, we know that an upper bound is given by:

$$f(x_k) - p_* \leq \frac{\beta}{2}\tilde{\gamma}^{2k} \cdot \|x_0 - x_*\|^2$$

Thm: Gerschgorin

Let A be a complex $n \times n$ matrix with entries $a_{i,j}$. Let R_i be the sum of the absolute values of the non-diagonal entries in the i-th row:

$$R_i := \sum_{j \neq i} |a_{ij}|$$

Let $D(a_{ii}, R_i) \subset \mathbb{C}$ be a closed disc centered at a_{ii} with radius R_i. Then every eigenvalue of A lies withing at least one of the Gershgorin discs.

This is used in the demonstration. Suppose $A \in \mathbb{S}_+^n$. Then all the eigenvalues are real and positive. For matrices with specific entries, e.g. tri-diagonal matrices, this can become handy as the sum R_i is easy to compute.

Theorem: Nemirovski-Yudin, Westerov

There exists:
$f : \ell_2 \to \mathbb{R}$ of type (α, β) such that for any method ϕ_i satisfying: $x_{k+1} = \text{span}(x_0, \dots, \nabla f(x_k))$

$$f(x_k) - p_* \geq \frac{\alpha}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k} \cdot \|x_0 - x_*\|^2 \qquad \kappa = \frac{\beta}{\alpha}$$

Conclusion: We see there's a large gap when $k \to \infty$:

$$\exp\left(\frac{-4k}{\sqrt{\kappa}}\right) \approx \frac{\alpha}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k} \leq \frac{\text{OC}_k(\mathcal{F})}{\|x_0 - x_*\|_2^2} \leq \frac{\beta}{2}\left(\frac{\kappa-1}{\kappa+1}\right)^{2k} \approx \exp\left(\frac{-4k}{\kappa}\right)$$

There's not much we can to to improve from the left but there exists a method which can improve the upper bound.

Nesterov Accelarated Gradiant Descent (AGD)

The algorithm looks like:

$$\begin{aligned} x_0 &= y_0 \\ y_{k+1} &= x_k - \frac{1}{\beta}\nabla f(x_K) \\ x_{k+1} &= (1+\delta)\cdot y_{k+1} - \delta \cdot y_k \qquad \delta = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \end{aligned}$$

Thm: Nesterov

For f of type (α, β), AGD satisfies:

$$f(x_k) - p_* \leq \frac{\alpha+\beta}{2}\exp\left(-\frac{k}{\sqrt{\kappa}}\right) \cdot \|x_0 - x_*\|_2^2$$

5.4 Smooth but not strongly convex

Direct consequences of removing the α strong convexity hypethesis:

- x^* is not necc. unique.
- This implies we can only talk about convergence in function value: $f(x_k) \to p_*$.
- No longer use bouds that contains $\frac{1}{\alpha}$
- The algorithms become much slower.

Thm: SD convergence

Let f be β-smooth. Then SD with stepsize $\frac{1}{\beta}$ satisfies:

$$f(x_k) - p_* \leq \frac{2\beta}{k} \cdot \|x_0 - x_*\|_2^2 \quad \text{with any} \quad x_* \ : \ f(x_*) = p_*$$

Rem: The convergence rate is algebraic: $\mathcal{O}(\frac{1}{k})$, instead of exponential. Meaning much slower.

Thm: Lower bound to oracle complexity

There exists $f : \mathbb{R}^n \to \mathbb{R}$ convex and β-smooth such that any method satisfying our conditions ($x_{k+1} \in \text{span}(x_0, \dots, \nabla f(x_0), \dots)$) also satisfies:

$$\min_{1 \leq \ell \leq k} f(x_\ell) - p_* \geq \frac{3\beta}{32}\frac{\|x_0 - x_*\|^2}{(k+2)^2}\left(\approx \mathcal{O}\left(\frac{\beta}{k^2}\right)\right)$$

Rem: This also leaves a gap in the bound as SD $\approx \frac{1}{k}$ and lower bound $\approx \frac{1}{k^2}$. Turns out that Nesterov also found a way to reduce the upper bound using a variation of ASD.

5.5 Subgradients

Let f no longer be smooth.

Def: Subgradient

Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$. We call $g \in \mathbb{R}^n$ a subgradient of f at $x \in \Omega$ if $\forall y \in \Omega$:

$$f(y) \geq f(x) + g^T(y - x)$$

Thm: Subgrad

Let $\Omega \subset \mathbb{R}^n$ be convex. Let $f : \Omega \to \mathbb{R}$

1. If $\forall x \in \Omega$: $\partial f(x) \neq \varnothing$ then f is convex.
2. If f is convex, then $\forall x \in \text{int } \Omega$: $\partial f(x) \neq \varnothing$

Thm: Supporting Hyperplane

Let $S \subset \mathbb{R}^n$ be convex and $x \in \partial S$. Then $\exists w \in \mathbb{R}^n,\ w \neq 0$ such that $\forall y \in S \quad : \quad w^T y \leq w^T x$

5.6 Optimality conditions

Thm: Optim cond. non smooth

Let f be convex. Then x^* is a global min of f iff $\vec{0} \in \partial f(x^*)$

Def: Normal cone to Ω at x:

Let $\Omega \subset \mathbb{R}^n$ be convex and closed:

$$N_\Omega(x) := \{v \in \mathbb{R}^n \mid v^T(y - x) \leq 0 \quad \forall y \in \Omega\}$$

Rem: On a smooth part of $\partial\Omega$, $N_\Omega(x)$ is the orthogonal complement of the tangent space $T_\Omega(x)$ in $\mathbb{R}^n$.

Thm: Opt. cond. constrained case

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $\Omega \subset \mathbb{R}^n$ be convex and non empty. Then x^* is a global min of $\min_{x \in \Omega} f(x)$ iff $x^* \in \Omega$ and:

$$\exists g \in \partial f(x^*) \quad \text{s.t.} \quad -g \in N_\Omega(x^*)$$

5.7 Projected Gradiant Descent

We want to solve $\min f(x)$ s.t. $x \in \Omega$. Steepest descent used something along the lines of: $x_1 = x_0 - t\nabla f(x_0)$. Problems:

- $\nabla f(x_0)$ isn't always defined, let's replace it with some $g \in \partial f(x_0)$.
- x_1 could escape our constain Ω. We can project it back into Ω: $\tilde{x_1} := x_0 - t \cdot g \rightsquigarrow x_1 = P_\Omega(\tilde{x_1})$ Where the projection is correctly define (See HW): $P_C(x) := \arg\min_{y \in \Omega} \|x - y\|_2^2$.

Algo: Projected GD

$$x_0 \in \Omega \qquad y_{k+1} = x_k - t_k \cdot g_k \quad g_k \in \partial f(x_k) \qquad x_{k+1} = P_\Omega(y_{k+1})$$

Thm: PGD convergence rate

Assuming:

A) Ω is connex, compact and non-empty:

- If $x \in \Omega$, $\|x\| \leq R \in \mathbb{R}$. Other words: bounded
- $P_\Omega(x)$ exists.

B) f is convex with bounded subgradiants: $\forall x \in \mathbb{R}^n \qquad \|g\| \leq L \quad \forall g \in \partial f(x)$. This mean f is L lipschitz:

$$|f(y) - f(x)| \leq |g^T(x-y)| \leq L\|x-y\|$$

Then projected DG with stepsize $t_k := \dfrac{R}{L\sqrt{k}}$ satisfies:

$$f\left(\frac{1}{k}\sum_{s=1}^{k} x_k\right) - p_* \leq \frac{RL}{\sqrt{k}}$$

Note:

- t_k is not constant, as $k \to \infty$, $t_k \to 0$
- We evaluate an average of lasts points.
- Convergence is very slow $\sim \dfrac{1}{\sqrt{k}}$

In some sense there's no better method. In fact one can show that there exists a function f for which any PGD will have error bonded $\geq \frac{1}{\sqrt{k}}$.