A Finite-Dimensional Geometric-Chain Lower Bound for Gradient-Span Methods

1 Statement of the result

Let $0 < \mu < L$, and define

$$\kappa = \frac{L}{\mu} > 1, \qquad q = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$

Since $\kappa > 1$, we have $0 < q < 1$.

We consider first-order methods initialized at $x_0 = 0$ and satisfying the *gradient-span condition*

$$x_{k+1} \in \operatorname{span}\{\nabla f(x_0), \dots, \nabla f(x_k)\}. \tag{1}$$

Thus x_T is the point produced after gradients have been evaluated at $x_0, \dots, x_{T-1}$.

Proposition 1. *Let $T \geq 1$ and $R > 0$. There exists a quadratic function $f : \mathbb{R}^{2T} \to \mathbb{R}$ such that*

$$\mu I \preceq \nabla^2 f \preceq LI, \qquad \|x_0 - x^\star\| = R,$$

and every method satisfying (1) *obeys*

$$f(x_T) - f(x^\star) \geq \frac{\mu R^2}{4} \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2T}.$$

Here $x^\star$ denotes the unique minimizer of f.

The proof has three ingredients:

1. the minimizer is the geometric sequence $(\alpha q, \alpha q^2, \dots, \alpha q^n)$;
2. the tridiagonal Hessian reveals at most one new coordinate per gradient evaluation;
3. after T evaluations, the geometric tail beyond coordinate T gives a lower bound on the distance to the minimizer.

2 Construction of the quadratic

We first work in an arbitrary dimension $n > T$. Define

$$B_n = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 - q \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

For every $x \in \mathbb{R}^n$, direct expansion gives

$$x^\top B_n x = x_1^2 + \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 + (1 - q) x_n^2. \tag{2}$$

Indeed, the coefficient of x_1^2 is $1 + 1 = 2$, the coefficient of each x_i^2, $2 \leq i \leq n - 1$, is $1 + 1 = 2$, and the coefficient of x_n^2 is $1 + (1 - q) = 2 - q$. The cross terms are $-2x_i x_{i+1}$, corresponding to off-diagonal entries -1.

Let $\alpha > 0$ be a scaling parameter and define

$$f(x) = \frac{\mu}{2} \|x\|^2 + \frac{\mu(\kappa - 1)}{8} \left(x^\top B_n x - 2\alpha x_1 \right). \tag{3}$$

The gradient and Hessian are

$$\nabla f(x) = \mu x + \frac{\mu(\kappa-1)}{4}\big(B_n x - \alpha e_1\big), \tag{4}$$

$$\nabla^2 f = \mu I + \frac{\mu(\kappa-1)}{4} B_n. \tag{5}$$

Strong convexity and smoothness

All terms on the right-hand side of (2) are nonnegative, so

$$x^\top B_n x \geq 0.$$

Therefore $B_n \succeq 0$.

For the upper bound, use

$$(x_i - x_{i+1})^2 \leq 2x_i^2 + 2x_{i+1}^2 \quad \text{and} \quad 1 - q \leq 1.$$

It follows that

$$\begin{aligned} x^\top B_n x &\leq x_1^2 + 2\sum_{i=1}^{n-1}(x_i^2 + x_{i+1}^2) + x_n^2 \\ &= 3x_1^2 + 4\sum_{i=2}^{n-1} x_i^2 + 3x_n^2 \\ &\leq 4\|x\|^2. \end{aligned}$$

Thus

$$0 \preceq B_n \preceq 4I.$$

Applying these inequalities to (5) gives

$$\nabla^2 f \succeq \mu I$$

and

$$\nabla^2 f \preceq \mu I + \frac{\mu(\kappa-1)}{4}(4I) = \mu\kappa I = LI.$$

Hence f is μ-strongly convex and L-smooth. In particular, f has a unique minimizer.

3 The minimizer is exactly geometric

Introduce the abbreviation

$$a = \frac{\kappa - 1}{4} > 0.$$

The first-order optimality condition $\nabla f(x^\star) = 0$ becomes

$$(I + aB_n)x^\star = a\alpha e_1. \tag{6}$$

We first record the algebraic identity that determines q. Let $s = \sqrt{\kappa}$. Since $q = (s-1)/(s+1)$,

$$\begin{aligned} q + \frac{1}{q} &= \frac{s-1}{s+1} + \frac{s+1}{s-1} \\ &= \frac{(s-1)^2 + (s+1)^2}{s^2 - 1} \\ &= \frac{2(s^2+1)}{s^2-1} = \frac{2(\kappa+1)}{\kappa-1}. \end{aligned}$$

On the other hand,

$$2 + \frac{1}{a} = 2 + \frac{4}{\kappa - 1} = \frac{2(\kappa+1)}{\kappa-1}.$$

Consequently,

$$q + \frac{1}{q} = 2 + \frac{1}{a}, \qquad \text{or equivalently} \qquad 1 + 2a = a\left(q + \frac{1}{q}\right). \tag{7}$$

Multiplying the last identity by q also gives

$$(1 + 2a)q = a(1 + q^2). \tag{8}$$

We now claim that

$$x_i^\star = \alpha q^i, \qquad i = 1, \dots, n. \tag{9}$$

We verify every equation in (6).

For the first coordinate, the equation is

$$(1 + 2a)x_1^\star - ax_2^\star = a\alpha.$$

Substituting $x_1^\star = \alpha q$ and $x_2^\star = \alpha q^2$, the left-hand side is

$$\alpha\big((1 + 2a)q - aq^2\big) = \alpha\big(a(1 + q^2) - aq^2\big) = a\alpha,$$

where we used (8).

For an interior coordinate $2 \le i \le n - 1$, the equation is

$$-ax_{i-1}^\star + (1 + 2a)x_i^\star - ax_{i+1}^\star = 0.$$

Substitution gives

$$\begin{aligned} &- a\alpha q^{i-1} + (1 + 2a)\alpha q^i - a\alpha q^{i+1} \\ &\quad = \alpha q^{i-1}\big[-a + (1 + 2a)q - aq^2\big] = 0, \end{aligned}$$

again by (8).

Finally, because the last diagonal entry of B_n is $2 - q$, the last equation is

$$-ax_{n-1}^\star + \big(1 + a(2 - q)\big)x_n^\star = 0.$$

After substitution, its left-hand side becomes

$$\begin{aligned} &- a\alpha q^{n-1} + \big(1 + 2a - aq\big)\alpha q^n \\ &\quad = \alpha q^{n-1}\big[-a + (1 + 2a)q - aq^2\big] = 0. \end{aligned}$$

Thus (9) satisfies all the equations in (6). Since f is strongly convex, this solution is the unique minimizer.

The modified final entry $2 - q$ is precisely what makes the finite geometric sequence satisfy the last equation. With final entry 2, the boundary would perturb the geometric solution.

Because $x_0 = 0$, the initial squared distance is

$$\begin{aligned} \|x_0 - x^\star\|^2 &= \alpha^2 \sum_{i=1}^{n} q^{2i} \\ &= \alpha^2 \frac{q^2(1 - q^{2n})}{1 - q^2}. \end{aligned} \tag{10}$$

Given a desired radius $R > 0$, choose

$$\alpha = \frac{R}{\sqrt{\sum_{i=1}^{n} q^{2i}}} = \frac{R\sqrt{1 - q^2}}{q\sqrt{1 - q^{2n}}}. \tag{11}$$

Then $\|x_0 - x^\star\| = R$.

4 Only one new coordinate is revealed per iteration

For $0 \le k \le n$, define

$$E_k = \operatorname{span}\{e_1, \dots, e_k\}, \qquad E_0 = \{0\}.$$

The tridiagonal structure implies

$$B_n E_k \subseteq E_{k+1}. \tag{12}$$

To see this explicitly, suppose $x \in E_k$, so that $x_i = 0$ for $i > k$. The i-th coordinate of $B_n x$ depends only on x_{i-1}, x_i, x_{i+1}. If $i > k+1$, all three of these coordinates vanish. Therefore $(B_n x)_i = 0$ for $i > k+1$, proving (12).

Using the gradient formula (4), if $x \in E_k$, then

$$x \in E_k \subseteq E_{k+1}, \qquad B_n x \in E_{k+1}, \qquad e_1 \in E_{k+1}.$$

Consequently,

$$x \in E_k \quad \Longrightarrow \quad \nabla f(x) \in E_{k+1}. \tag{13}$$

We now prove by induction that

$$x_k \in E_k \qquad \text{for every } k \le T. \tag{14}$$

The base case holds because $x_0 = 0 \in E_0$. Suppose that $x_j \in E_j$ for $j = 0, \dots, k$. By (13),

$$\nabla f(x_j) \in E_{j+1} \subseteq E_{k+1}, \qquad j = 0, \dots, k.$$

The gradient-span condition therefore implies

$$x_{k+1} \in \operatorname{span}\{\nabla f(x_0), \dots, \nabla f(x_k)\} \subseteq E_{k+1}.$$

This completes the induction. In particular,

$$x_T \in E_T, \qquad \text{so} \qquad (x_T)_i = 0 \quad \text{for } i > T. \tag{15}$$

5 The hidden geometric tail

The definition of μ-strong convexity gives, for every x,

$$f(x) \ge f(x^\star) + \langle \nabla f(x^\star), x - x^\star \rangle + \frac{\mu}{2}\|x - x^\star\|^2.$$

Since $\nabla f(x^\star) = 0$, this implies

$$f(x_T) - f(x^\star) \ge \frac{\mu}{2}\|x_T - x^\star\|^2.$$

Because $(x_T)_i = 0$ for $i > T$,

$$\begin{aligned} \|x_T - x^\star\|^2 &= \sum_{i=1}^{T} \big((x_T)_i - x_i^\star\big)^2 + \sum_{i=T+1}^{n} (x_i^\star)^2 \\ &\ge \sum_{i=T+1}^{n} (x_i^\star)^2. \end{aligned}$$

Using $x_i^\star = \alpha q^i$, we obtain

$$f(x_T) - f(x^\star) \ge \frac{\mu\alpha^2}{2} \sum_{i=T+1}^{n} q^{2i}. \tag{16}$$

We express this in terms of R. The two relevant geometric sums are

$$\begin{aligned} \sum_{i=1}^{n} q^{2i} &= \frac{q^2(1 - q^{2n})}{1 - q^2}, \\ \sum_{i=T+1}^{n} q^{2i} &= \frac{q^{2(T+1)}(1 - q^{2(n-T)})}{1 - q^2}. \end{aligned}$$

Their ratio is

$$\frac{\sum_{i=T+1}^{n} q^{2i}}{\sum_{i=1}^{n} q^{2i}} = q^{2T} \frac{1 - q^{2(n-T)}}{1 - q^{2n}}.$$

Since $R^2 = \alpha^2 \sum_{i=1}^{n} q^{2i}$, inequality (16) becomes

$$f(x_T) - f(x^\star) \geq \frac{\mu R^2}{2} q^{2T} \frac{1 - q^{2(n-T)}}{1 - q^{2n}}. \tag{17}$$

To prove Proposition 1, set $n = 2T$. Then

$$\begin{aligned} \frac{1 - q^{2(n-T)}}{1 - q^{2n}} &= \frac{1 - q^{2T}}{1 - q^{4T}} \\ &= \frac{1 - q^{2T}}{(1 - q^{2T})(1 + q^{2T})} \\ &= \frac{1}{1 + q^{2T}} \geq \frac{1}{2}, \end{aligned}$$

because $0 < q^{2T} < 1$. Substituting this into (17) yields

$$f(x_T) - f(x^\star) \geq \frac{\mu R^2}{4} q^{2T} = \frac{\mu R^2}{4} \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2T}.$$

Together with (11), this explicitly constructs a function on $\mathbb{R}^{2T}$ with $\|x_0 - x^\star\| = R$, completing the proof.

Why the construction depends on T. The dimension $n = 2T$ is chosen for a lower bound at a prescribed finite horizon T. A fixed finite-dimensional quadratic cannot produce this obstruction indefinitely: after sufficiently many iterations, Krylov-space methods may explore the entire space. The finite chain therefore replaces the infinite-dimensional geometric tail by a tail that remains sufficiently long up to iteration T.