Introduction to lower bounds in optimization

Samuel Vaiter

June 12, 2025

This content is available as a webpage at https://samuelvaiter.com/intro-to-lower-bounds/.

Most of the bounds that are described in the optimization litterature are *upper* bounds of the form $\|x^{(t)} - x^\star\| \leq \alpha(t)$ or $f(x^{(t)}) - f(x^\star) \leq \alpha(t)$. But what about findings the reverse-side inequality $\|x^{(t)} - x^\star\| \geq \beta(t)$? Said otherwise, what can we achieve with a "gradient-descent-like" algorithm?

To formalize this notion, we consider algorithm, here sequences $(x^{(t)})_{t\geq 0}$, that build upon the previous iterates with only access to a first-order oracle:

Assumption 1 (First-order method)**.** We assume that a first-order method is given by a sequence $x^{(t)}$ such that

$$x^{(t)} \in x^{(0)} + \operatorname{Span}\{\nabla f(x^{(0)}), \dots, \nabla f(x^{(t-1)})\}.$$

We shall note that one can thinks of a more general way to define first-order methods, but for the sake of the results we aim to prove, such level of generality is enough.

With this assumption in mind, how to design a function adversarial to these type of schemes? The idea is to find a function such that the gradient at step $t-1$ gives minimal information, *i.e.*, it has a minimal nonzero partial derivatives. A way to define such function is to "stack" quadratic functions with increasing dependencies between variables:

$$f_k^{L,\mu}(x) = \frac{L-\mu}{8}\left((x_1-1)^2 + \sum_{i=1}^{k-1}(x_{i+1}-x_i)^2 + x_k^2\right) + \frac{\mu}{2}\|x\|^2, \tag{1}$$

where $0 \leq \mu < L$ and $0 \leq k \leq d$.

$$\frac{\partial f_k^{L,\mu}}{\partial x_i}(x) = \mu x_i + \frac{L-\mu}{4}\begin{cases} -x_2 + 2x_1 - 1 & \text{if } i = 1 \\ -x_{i+1} + 2x_i - x_{i-1} & \text{if } 2 \leq i < k \\ 2x_k - x_{k-1} & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases}$$

Set $f = f_k^{L,\mu}$. Observe that if we start from $x^{(0)} = 0$, then

$$x^{(1)} = x^{(0)} - \eta \nabla f(0) = -\eta \frac{L-\mu}{4} e_1 \in \mathbb{R} e_1,$$

that is only the first coordinate is updated after one iteration. What happens now that we have access to $x^{(0)}$, $\nabla f_k(x^{(0)})$ and $\nabla f_k(x^{(1)})$? An algorithm satisfying Assumption 1, we look at

$$x^{(2)} = x^0 + \alpha \nabla f(x^{(0)}) + \beta \nabla f(x^{(1)}).$$

One can check that for any (α, β), $x^{(2)} \in \mathbb{R}e_1 + \mathbb{R}e_2$, and by an easy induction, we have $x^{(t)} \in \sum_{k=1}^{t} \mathbb{R} e_k$: any first-order methods will only be able to update at most one new coordinate at each iteration. We are going to prove the following result.

Theorem 1 (Lower-bound for smooth convex optimization)**.** *For any $d \geq 2$, $x^{(0)} \in \mathbb{R}^d$, $L > 0$, $t \leq (n-1)/2$, there exists a convex function f that is C^∞ and L-smooth such that any sequences satisfying Assumption 1 is such that*

$$f(x^{(t)}) - f(x^\star) \geq \frac{3L\|x^{(0)} - x^\star\|^2}{32(t+1)^2} \tag{2}$$

$$\tag{3}$$

where $x^\star$ is a minimizer of f.

Remark that the rate $1/t^2$ is not achieved by the gradient descent[1]! We also have a lower bound for the class of strongly convex functions.

Theorem 2 (Lower-bound for smooth strongly convex optimization)**.** *For any $d \geq 2$, $x^{(0)} \in \mathbb{R}^d$, $L > 0$, there exists a μ-strongly convex function f that is C^∞ and L-smooth such that any sequences satisfying Assumption 1 is such that for all $t < (n-1)/2$, we have*

$$\|x^{(t)} - x^\star\|^2 \geq \frac{1}{8} \left(\frac{\sqrt{K_f} - 1}{\sqrt{K_f} + 1} \right)^{2t} \|x^{(0)} - x^\star\|^2, \tag{4}$$

$$f(x^{(t)}) - f(x^\star) \geq \frac{\mu}{16} \left(\frac{\sqrt{K_f} - 1}{\sqrt{K_f} + 1} \right)^{2t} \|x^{(0)} - x^\star\|^2. \tag{5}$$

where $x^\star$ is the unique minimizer of f.

[1] There exist algorithms that achieve it, in particular Nesterov's acceleration method.

Note that it is common in the litterature to see Theorem 2 without the factor $\frac{1}{8}$. This is due to an artefact of proof since we prove this result in the finite dimensional case whereas Nesterov (2018) works in the infinite dimensional space $\ell^2(\mathbb{N})$. Before proving these important results due to (Nemirovski and Yudin, 1983), we are going to prove several lemmas.

Lemma 1 (Minimizers of f_k)**.** *Let $d \geq 2$, $L > 0$, $\mu \geq 0$, then $f_k^{L,\mu}$ defined in* (1) *is a μ-strongly convex (eventually convex if $\mu = 0$) C^∞-function such that its gradient is L-Lipschitz.*

If $\mu = 0$, it has a unique minimizer $x^{k,\star}$ satisfying

$$x_i^{k,\star} = \begin{cases} 1 - \frac{i}{k+1} & \text{if } 1 \leq i \leq k \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad f_k^{L,0}(x^{k,\star}) = \frac{L}{8(k+1)}.$$

If $\mu > 0$, it has a unique minimizer $x^{k,\star}$ satisfying

$$x_i^{k,\star} = \frac{s^{2(k+1)}}{s^{2(k+1)} - 1} s^{-i} + \frac{1}{1 - s^{2(k+1)}} s^i,$$

for $1 \leq i \leq k$, and $x_i^{k,\star} = 0$ for $i > k$, where $s = \frac{\sqrt{K_f}+1}{\sqrt{K_f}-1}$.

Proof. We drop the exponents L, μ in the definition of $f_k = f_k^{L,\mu}$. The function f_k being a quadratic form, it is C^∞ and its partial derivatives read

$$\frac{\partial f_k}{\partial x_i \partial x_j}(x) = \mu 1_{\{i=j\}} + \frac{L-\mu}{4} \begin{cases} 2 & \text{if } i = j \leq k \\ -1 & \text{if } j = i - 1 \text{ and } 1 < i \leq k \\ -1 & \text{if } j = i + 1 \text{ and } 1 \leq i < k \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the Hessian matrix is given (for any $x \in \mathbb{R}^d$) by

$$\nabla^2 f_k(x) = \mu \operatorname{Id}_d + \frac{L-\mu}{4} L_k,$$

where L_k is a (thresholded) discrete Laplacian operator with Dirichlet boundary conditions that is tridiagonal

$$L_k = \left(\begin{array}{ccccc|c} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & & & \\ & -1 & \ddots & \ddots & & 0_{k,d-k} \\ & & \ddots & \ddots & -1 & \\ & & & -1 & 2 & \\ \hline & & 0_{d-k,k} & & & 0_{d-k,d-k} \end{array}\right).$$

Observe that we have (since f_k is a quadratic form)

$$f_k(x) = \frac{1}{2}\langle \nabla^2 f_k(x)x, x\rangle - \frac{L-\mu}{4}x_1 + \frac{L-\mu}{8}.$$

Note that:

1. The Hessian is definite (resp. semi-definite) positive if $\mu > 0$ (resp. $\mu = 0$). Indeed,

$$\langle \nabla^2 f_k(x)h, h\rangle = \mu\|h\|^2 + \frac{L-\mu}{4}\langle L_k h, h\rangle.$$

   Since $\langle L_k h, h\rangle = h_1^2 + \sum_{i=1}^{k}(h_{i+1} - h_i)^2 + h_k^2 \geq 0$ for any h, the result follows depending on the value of μ.

2. Since $(a-b)^2 \leq 2a^2 + 2b^2$, we have

$$h_1^2 + \sum_{i=1}^{k}(h_{i+1} - h_i)^2 + h_k^2 \leq h_1^2 + \sum_{i=1}^{k}(2h_{i+1}^2 + 2h_i^2) + h_k^2 \leq 4\sum_{i=1}^{k} h_i^2 \leq 4\sum_{i=1}^{d} h_i^2 = 4\|h\|^2.$$

   Hence,

$$\langle \nabla^2 f_k(x)h, h\rangle \leq \mu\|h\|^2 + (L-\mu)\|h\|^2 = L\|h\|^2.$$

Thus, we have $\mu\,\mathrm{Id} \preceq \nabla^2 f_k(x) \preceq L\,\mathrm{Id}$.

Let us characterize the (unique) solution $x^{k,\star}$ of the minimization of f_k over $\mathbb{R}^d$. We aim to solve $\nabla f_k(x^{k,\star}) = 0$ to find a critical point (which will be a minimum since we just proved that the Hessian is at least semidefinite positive), that is

$$\mu x^{k,\star} + \frac{L-\mu}{4}L_k x^{k,\star} - \frac{L-\mu}{4}e_1 = 0.$$

Projecting this relation on each coordinate $2 \leq i \leq k-1$, we get that

$$-x_{i-1}^{k,\star} + 2x_i^{k,\star} - x_{i-1}^{k,\star} = -\frac{4\mu}{L-\mu}x_i^{k,\star},$$

which leads to

$$x_i^{k,\star} = \frac{1}{2}\frac{L+\mu}{L-\mu}(x_{i+1}^{k,\star} + x_{i-1}^{k,\star}).$$

Similarly, we have

$$x_1^{k,\star} = \frac{1}{2}\frac{L+\mu}{L-\mu}(x_2^{k,\star} + 1) \quad \text{and} \quad x_k^{k,\star} = \frac{1}{2}\frac{L+\mu}{L-\mu}x_{k-1}^{k,\star}.$$

Consider $y_0, \ldots, y_{k+1}$ defined by $y_i = x_i^{k,\star}$ for $1 \le i \le k$ and $y_0 = 1$ and $y_{k+1} = 0$. We have the relation

$$y_i = \alpha(y_{i+1} + y_{i-1}) \quad \text{where} \quad \alpha = \frac{1}{2}\frac{L+\mu}{L-\mu} > 0.$$

We can rewrite it as the second-order linear recursion $y_{i+2} - \alpha^{-1} y_{i+1} + y_i = 0$. The associated trinom is $P = X^2 - \alpha^{-1}X + 1 \in \mathbb{R}[X]$ whose discriminant is given by

$$\Delta = (-\alpha^{-1})^2 - 4 = 16\frac{L\mu}{(L-\mu)^2}.$$

We distinguish two cases:

1. If $\mu = 0$, then the unique root is given by $r = 1$.
2. If $\mu > 0$, then the roots are given by

$$\begin{aligned} r &= \frac{1}{2}(\alpha^{-1} - \sqrt{\Delta}) = \frac{\sqrt{\frac{L}{\mu}} - 1}{\sqrt{\frac{L}{\mu}} + 1} = \frac{\sqrt{K_f} - 1}{\sqrt{K_f} + 1} \\ s &= \frac{1}{2}(\alpha^{-1} + \sqrt{\Delta}) = \frac{\sqrt{K_f} + 1}{\sqrt{K_f} - 1} = \frac{1}{r}. \end{aligned}$$

Case $\mu = 0$. We have the affine relation $y_i = (a + bi)r$ with constraints $y_0 = a = 1$ and $y_{k+1} = a + b(k+1) = 0$. In turn, we have $y_i = 1 - \frac{i}{k+1}$ and thus

$$x_i^{k,\star} = \begin{cases} 1 - \frac{i}{k+1} & \text{if } 1 \le i \le k \\ 0 & \text{otherwise.} \end{cases}$$

The associated optimal value is given by

$$f_k(x^{k,\star}) = \frac{L}{8}\left(\left(-\frac{1}{k+1}\right) + \sum_{i=1}^{k-1} \frac{1}{(k+1)^2} + \left(1 - \frac{k}{k+1}\right)^2\right) = \frac{L}{8}\frac{k+1}{(k+1)^2} = \frac{L}{8}\frac{1}{k+1}.$$

Case $\mu > 0$. The solution can be written as

$$y_i = ar^i + bs^i \quad \text{with} \quad \begin{cases} a + b & = 1 \\ ar^{k+1} + bs^{k+1} & = 0. \end{cases}$$

Thus, we have $b = 1 - a$, hence $\frac{a}{a-1} = s^{2(k+1)} > 0$, and in turn we have

$$a = \frac{s^{2(k+1)}}{s^{2(k+1)} - 1} \quad \text{and} \quad b = \frac{1}{1 - s^{2(k+1)}}.$$

Hence,

$$y_i = \frac{s^{2(k+1)}}{s^{2(k+1)} - 1} s^{-i} + \frac{1}{1 - s^{2(k+1)}} s^i.$$

□

We now turns to the proofs of Theorem 1 and Theorem 2.

Proof of Theorem 1. We restrict our attention the the case where $x_0 = 0$ w.l.o.g. Indeed, if $x_0 \neq 0$, we can set $x \mapsto \tilde{f}(x) = f(x + x_0)$ and the following proof carry on. Let $d = 2k + 1$ and set $f = f_{2k+1}^{L,0}$.

Remark that

$$f(x^{(t)}) = f_{2k+1}^{L,0}(x^{(t)}) = f_t^{L,0}(x^{(t)}) \geq f_t^{\star}.$$

Using Lemma 1, we have on one hand that $f_t^{\star} = \frac{L}{8(t+1)}$, and then

$$f(x^{(t)}) - f(x^{\star}) = \frac{L}{8(k+1)} - \frac{L}{8\dot{2}(k+1)} = \frac{L}{16(k+1)}.$$

On the other hand,

$$\begin{aligned}
&\|x^{2k+1,\star} - x_0\|^2 = \|x^{2k+1,\star}\|^2 = \sum_{i=1}^{2k+1} (x^{2k+1,\star})_i^2 \\
&= \sum_{i=1}^{2k+1} \left(1 - \frac{i}{2(k+1)}\right)^2 \\
&= \sum_{i=1}^{2k+1} 1 - \frac{2}{2(k+1)} \sum_{i=1}^{2k+1} i + \frac{1}{4(k+1)^2} \sum_{i=1}^{2k+1} i^2 \\
&= (2k+1) - \frac{1}{k+1} \frac{2(k+1)(2k+1)}{2} + \frac{1}{4(k+1)^2} \frac{2(k+1)(4k+3)(2k+1)}{6} \\
&= \frac{1}{3} \frac{(2k+1)(4k+3)}{4(k+1)} \\
&\leq \frac{2k+1}{3} \leq \frac{2}{3}(k+1).
\end{aligned}$$

Thus,

$$\frac{f(x^{(t)}) - f(x^{\star})}{\|x^{2k+1,\star} - x_0\|^2} \geq \frac{\frac{L}{16(k+1)}}{\frac{2}{3}(k+1)} = \frac{3L}{32(k+1)^2},$$

that proves (2).

□

Proof of Theorem 2. The proof follows the same strategy as before, but we start with a bound on the iterates instead of the objective function. Assume that $x^{(0)} = 0$, otherwise let $\tilde{f} = f(\cdot + x_0)$. Consider $d = 2k+1$ and $f = f_{2k+1}^{L,\mu}$. We rewrite the coordinate of $x^{2k+1,\star}$ as

$$x_i^{2k+1,\star} = \frac{s^{4(k+1)}}{s^{4(k+1)} - 1} s^{-i} + \frac{1}{1 - s^{4(k+1)}} s^i = s^{-i} \left(1 - \frac{s^{2i} - 1}{s^{4(k+1)} - 1}\right).$$

On one hand, we have:

$$\|x^{(0)} - x^{2k+1,\star}\|^2 = \sum_{i=1}^{2k+1} (x_i^{2k+1,\star})^2 = \sum_{i=1}^{2k+1} s^{-2i} \left(1 - \frac{s^{2i} - 1}{s^{4(k+1)} - 1}\right)^2 \le \sum_{i=1}^{2k+1} s^{-2i},$$

where we used that for all $1 \le i \le 2k+1$, we have

$$0 \le 1 - \frac{s^{2i} - 1}{s^{4(k+1)} - 1} \le 1.$$

Bounding the tail of the geometric sums, we obtain

$$\|x^{(0)} - x^{2k+1,\star}\|^2 \le 2 \sum_{i=1}^{k+1} s^{-2i}. \tag{6}$$

On the other hand, observe that for $t < k+1$, one has $x^{(t)} \in \mathbb{R}^{k,2k+1}$, thus

$$\|x^{(t)} - x^{2k+1,\star}\|^2 \ge \sum_{i=k+1}^{2k+1} (x_i^{2k+1,\star})^2 = \sum_{i=k+1}^{2k+1} s^{-2i} \left(1 - \frac{s^{2i} - 1}{s^{4(k+1)} - 1}\right)^2.$$

Since $s > 0$, we have

$$\frac{1 - s^{2i}}{s^{4(k+1)} - 1} \le \frac{1 - s^{2(k+1)}}{s^{4(k+1)} - 1},$$

and in turn,

$$1 \ge \left(1 - \frac{s^{2i} - 1}{s^{4(k+1)} - 1}\right)^2 \ge \left(1 - \frac{s^{2(k+1)} - 1}{s^{4(k+1)} - 1}\right)^2 \ge 0.$$

Thus,

$$\begin{aligned} \|x^{(k)} - x^{2k+1,\star}\|^2 &\ge \left(1 - \frac{s^{2(k+1)} - 1}{s^{4(k+1)} - 1}\right)^2 \sum_{i=k+1}^{2k+1} s^{-2i} \\ &= \left(1 - \frac{s^{2(k+1)} - 1}{s^{4(k+1)} - 1}\right)^2 s^{-2k} \sum_{i=1}^{k+1} s^{-2i}. \end{aligned} \tag{7}$$

Observe that

$$\left(1 - \frac{s^{2(k+1)} - 1}{s^{4(k+1)} - 1}\right)^2 \geq \frac{1}{4}.$$

Combining it with (6) and (7), we have

$$\|x^{(t)} - x^{2k+1,\star}\|^2 \geq \frac{1}{8} s^{-2k} \|x^{(0)} - x^{2k+1,\star}\|^2 \geq \frac{1}{8} \left(\frac{\sqrt{K_f} - 1}{\sqrt{K_f} + 1}\right)^{2t} \|x^{(0)} - x^{2t+1,\star}\|^2,$$

proving (4). The value bound (5) is obtained by applying the following inequality

$$f(x) \geq f(x^\star) + \frac{\mu}{2}\|x - x^\star\|^2,$$

to f. $\square$

More refined versions of these lower bounds, which provide tighter constants, can be found in Drori and Taylor (2022). However, these improvements come at the price of significantly more involved and technical proofs.

References

Drori, Yoel and Adrien Taylor (Feb. 1, 2022). “On the Oracle Complexity of Smooth Strongly Convex Minimization”. In: *Journal of Complexity* 68, p. 101590.

Nemirovski, Arkadi and David Yudin (1983). *Problem Complexity and Method Efficiency in Optimization*. Wiley. 388 pp.

Nesterov, Yurii (2018). *Lectures on Convex Optimization*. Vol. 137. Springer.