Supplemental Material for 04/23 Book Reading Seminar

Hiroki Hamaguchi

April 23, 2025

Abstract

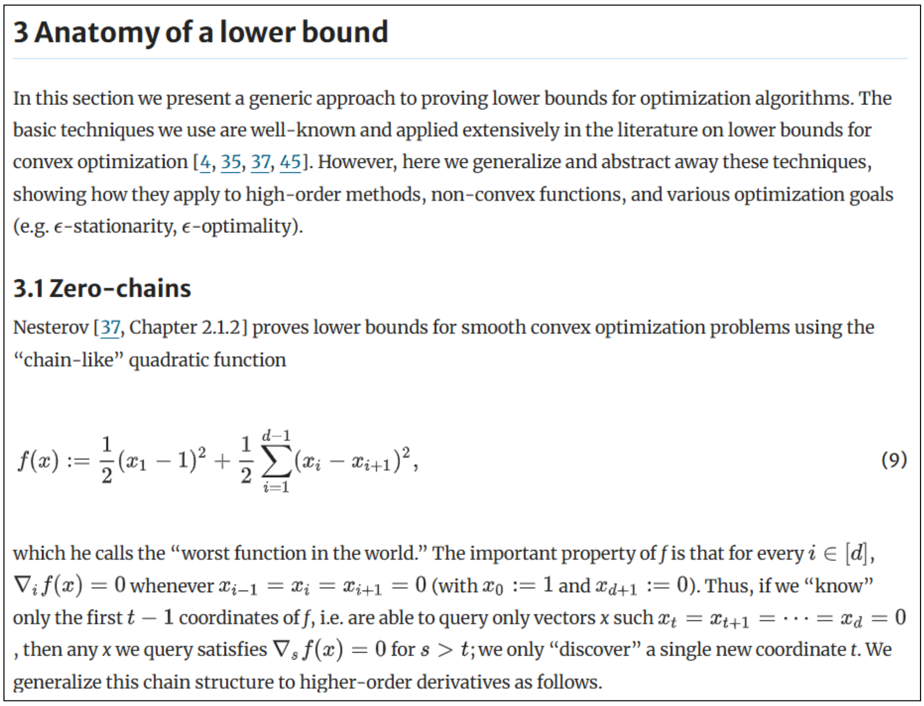
This is the supplemental material for the book reading seminar. Please refer to the textbook and the whiteboard for the main content. If necessary, please also refer to the repository.

2.2

2.2.1

Lemma 1.19: For any $f \in C^2(\mathbb{R}^n)$, and any $x, y \in \mathbb{R}^n$, there exists $z \in [x, y]$ s.t.

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x)^\top \nabla^2 f(z)(y - x).$$



3 Anatomy of a lower bound

In this section we present a generic approach to proving lower bounds for optimization algorithms. The basic techniques we use are well-known and applied extensively in the literature on lower bounds for convex optimization [4, 35, 37, 45]. However, here we generalize and abstract away these techniques, showing how they apply to high-order methods, non-convex functions, and various optimization goals (e.g. ϵ-stationarity, ϵ-optimality).

3.1 Zero-chains

Nesterov [37, Chapter 2.1.2] proves lower bounds for smooth convex optimization problems using the "chain-like" quadratic function

$$f(x) := \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}\sum_{i=1}^{d-1}(x_i - x_{i+1})^2, \tag{9}$$

which he calls the "worst function in the world." The important property of f is that for every $i \in [d]$, $\nabla_i f(x) = 0$ whenever $x_{i-1} = x_i = x_{i+1} = 0$ (with $x_0 := 1$ and $x_{d+1} := 0$). Thus, if we "know" only the first $t - 1$ coordinates of f, i.e. are able to query only vectors x such $x_t = x_{t+1} = \cdots = x_d = 0$, then any x we query satisfies $\nabla_s f(x) = 0$ for $s > t$; we only "discover" a single new coordinate t. We generalize this chain structure to higher-order derivatives as follows.

Figure 1: [15] から引用

参考文献の [15] は、Math. Program. の 2020 年の論文。

2.3

$A \succ O \iff \exists$ 直交行列 Q と対角行列 D s.t. $A = QDQ^\top$

2.4

Ex.2.15

一般に、

$$
\begin{aligned}
f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|y - x\|_2^2 \\
f(x) &\geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\mu}{2}\|y - x\|_2^2 \\
0 &\geq \langle \nabla f(x) - \nabla f(y), y - x \rangle + \mu\|y - x\|_2^2
\end{aligned}
$$

なので、

$$
\begin{aligned}
\langle \nabla f(x) - \nabla f(y), x - y \rangle &\geq \mu\|y - x\|_2^2 \\
\langle \nabla f(x), x - x^* \rangle &\geq \mu\|x - x^*\|_2^2
\end{aligned}
$$

である。

Lemma 2.10 より、

$$
\begin{aligned}
\left\|x^{(k+1)} - x^*\right\|_2^2 &= \left\|x^{(k)} - x^*\right\|_2^2 - \frac{1}{L}\langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle + \frac{1}{L^2}\left\|\nabla f(x^{(k)})\right\|_2^2 \\
&= \left\|x^{(k)} - x^*\right\|_2^2 - \frac{1}{L}\langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle \\
&\quad - \frac{1}{L}\left(\langle \nabla f(x^{(k)}), x^{(k)} - x^* \rangle - \frac{1}{L}\left\|\nabla f(x^{(k)})\right\|_2^2\right) \\
&\leq \left\|x^{(k)} - x^*\right\|_2^2 - \frac{\mu}{L}\left\|x^{(k)} - x^*\right\|_2^2
\end{aligned}
$$

途中でも Lemma 2.10 の証明中の式を使っている。

Ex.2.16

特段の名前はついてない? 自信なし。

Ex.2.17

PL 不等式

Ex.2.18

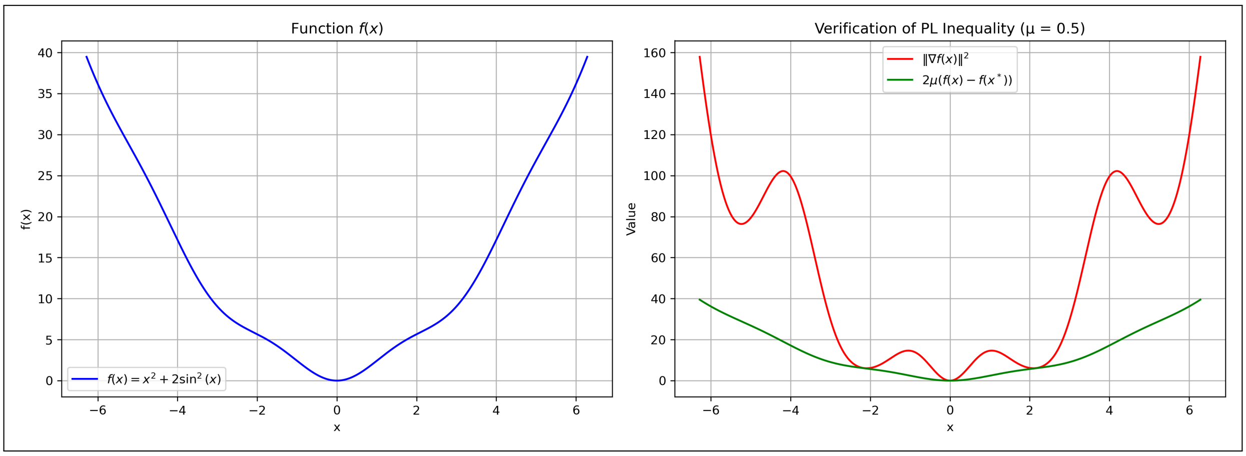


Figure 2: PL 不等式を満たすが非凸な例

2.5

直線探索

Ex.2.20

$$
\begin{aligned}
&-c_1 h\|\nabla f(x)\|_2^2 \\
&\geq f(x - h\nabla f(x)) - f(x) \\
&\geq \nabla f(x)^\top((x - h\nabla f(x)) - x) + \frac{\mu}{2}\|x - h\nabla f(x) - x\|_2^2 \\
&= -h\|\nabla f(x)\|_2^2 + \frac{\mu}{2}h^2\|\nabla f(x)\|_2^2
\end{aligned}
$$

$$
\therefore \quad -c_1 h \geq -h + \frac{\mu}{2}h^2 \iff \frac{2(1-c)}{\mu} \geq h
$$

$$
\begin{aligned}
&c_2\|\nabla f(x)\|_2^2 \\
&\geq \nabla f(x)^\top \nabla f(x - h\nabla f(x)) \\
&= \nabla f(x)^\top(\nabla f(x) - \nabla^2 f(z)(h\nabla f(x))) \\
&= \|\nabla f(x)\|_2^2 - h\nabla f(x)^\top \nabla^2 f(z)\nabla f(x)
\end{aligned}
$$

$$
\therefore c_2 \geq 1 - h\lambda_{\max}(\nabla^2 f(z)) \geq 1 - hL \iff h \leq \frac{1-c_2}{L}
$$

この証明は Taylor 展開と積分の平均値の定理を経由して考えないと誤りかも

2.6

Ex.2.23

不明。 私の頭が悪いだけかも知れません。
$h(y/\alpha)$ と y が 0 を中心にスケーリングされているのが謎。x が中心では?

Projected Gradient Descent / Proximal Gradient Descent

参考

Other assumptions

参考文献 [10] Iterative hard thresholding for compressed sensing
参考文献 [10] の参考文献 [1] Iterative Thresholding for Sparse Approximations

638 J Fourier Anal Appl (2008) 14: 629–654

3 An Iterative Algorithm for the M-Sparse Problem

3.1 Digest: The M-Sparse Algorithm

In this section we turn to the M-sparse problem

$$\min_{\mathbf{y}} \|\mathbf{x} - \Phi\mathbf{y}\|_2^2 \quad \text{subject to} \quad \|\mathbf{y}\|_0 \leq M, \tag{3.1}$$

and derive the following iterative algorithm

$$\mathbf{y}^{n+1} = H_M(\mathbf{y}^n + \Phi^H(\mathbf{x} - \Phi\mathbf{y}^n)), \tag{3.2}$$

where H_M is now a non-linear operator that only retains the M coefficients with the largest magnitude

$$H_M(y_i) \begin{cases} 0 & \text{if } |y_i| < \lambda_M^{0.5}(\mathbf{y}), \\ y_i & \text{if } |y_i| \geq \lambda_M^{0.5}(\mathbf{y}). \end{cases} \tag{3.3}$$

The threshold $\lambda_M^{0.5}(\mathbf{y})$ is set to the M^{th} largest absolute value of $\mathbf{y}^n + \Phi^H(\mathbf{x} - \Phi\mathbf{y}^n)$, if less than M values are non-zero we define $\lambda_M^{0.5}(\mathbf{y})$ to be the smallest absolute value of the non-zero coefficients. We call this algorithm the *M-sparse algorithm*.

Figure 3: Iterative Hard Thresholding

3

もしここまで到達してしまったらすいません、お詫びします。準備不足です。
質問とか大事な部分の深掘りとかで時間を使います。

Appendix

Listing 1: PL

```
import numpy as np
import matplotlib.pyplot as plt
import os

os.chdir(os.path.dirname(os.path.abspath(__file__)))

x = np.linspace(-2 * np.pi, 2 * np.pi, 1000)

f = x**2 + 2 * np.sin(x) ** 2
grad_f = 2 * x + 2 * np.sin(2 * x)

mu = 0.5
lhs = grad_f**2
rhs = 2 * mu * f

fig, axes = plt.subplots(1, 2, figsize=(14, 5))

axes[0].plot(x, f, label=r"$f(x) = x^2 + 2\sin^2(x)$", color="b
axes[0].set_title("Function $f(x)$")
axes[0].set_xlabel("x")
axes[0].set_ylabel("f(x)")
axes[0].legend()
axes[0].grid(True)

axes[1].plot(x, lhs, label=r"$\|\nabla f(x)\|^2$", color="red")
axes[1].plot(x, rhs, label=r"$2 \mu (f(x) - f(x^*))$", color="g
axes[1].set_title(f"Verification of PL Inequality μ ( = {mu})")
axes[1].set_xlabel("x")
axes[1].set_ylabel("Value")
axes[1].legend()
axes[1].grid(True)

plt.tight_layout()
# plt.show()
plt.savefig("PL.png", dpi=300, bbox_inches="tight")
```