Convergence rates for the Frank-Wolfe algorithm

Matteo Sammut

January 22, 2024

This project aims at solving the constrained optimization problem:

$$\min_{x \in C} f(x) \tag{1}$$

where $C \subseteq \mathbb{R}^d$ is a nonempty compact convex set, and $f : \mathbb{R}^d \to \mathbb{R}$ is a C^1 function, using the Frank-Wolfe algorithm. Our goal is to investigate the convergence rates provided in [1].

Question 1

Given a vector $\mathbf{a} \in \mathbb{R}^d$, we want to find a solution to the linear minimization problem

$$\min_{\mathbf{p} \in C} \langle a, p \rangle, \tag{2}$$

where C is C_2, C_∞, or $C_\mathcal{P}$.

(a) C_∞ and C_P are polytope. Also, in this question we determine the extreme points of those convex sets.
We prove that $\text{extr}(C_2) = \text{S}(0,1)$. We first prove that $\text{S}(0,1) \subseteq \text{extr}(C_2)$ by showing that $\text{S}(0,1)$ are exposed points which is a sufficient condition. That is, we show that $\forall x \in \text{S}(0,1), \exists v \in C_2 \setminus \{0\}$ such that :

$$x = \arg\max_{c \in C_2} \langle c, v \rangle$$

Let $x \in \text{S}(0,1), c \in C_2$.

$$\begin{aligned}
\langle c, x \rangle &\leq \|c\|_2 \|x\|_2 \quad \text{by Cauchy-Swhartz inequality} && (2)\\
&\leq \|x\|_2 \quad \text{as } c \in C_2 && (3)\\
&= 1 \quad \text{as } x \in \text{S}(0,1) && (4)
\end{aligned}$$

and $c = x$ achieves this bound.
To show that $\text{extr}(C_2) \subseteq \text{S}(0,1)$, we just have to see that if $x \in \text{extr}(C_2)$ such that $\|x\|_2 < 1$, then x belongs to the open line segment $]0, \frac{x}{\|x\|_2}[$ in C_2.
We prove that $\text{extr}(C_\infty) = \{x \in \mathbb{R}^d \,|\, \forall i \in \{1, \dots, d\}, |x_i| = 1\}$. Let $c \in C_\infty$ and $x \in \{x \in \mathbb{R}^d \,|\, \forall i \in \{1, \dots, d\}, |x_i| = 1\}$.

$$\begin{aligned}
\langle c, x \rangle = \sum_{i=1}^d c_i x_i &= \sum_{i=1}^d |c_i| \text{sign}(c_i) |x_i| \text{sign}(x_i) && (5)\\
&\leq \sum_{i=1}^d |vi| = d && (6)
\end{aligned}$$

and c achieves this bound if and only if $c_i = \text{sign}(x_i), \forall i \in \{1, \dots, d\}$. Using the same argument than for C_2, we have $\{x \in \mathbb{R}^d \,|\, \forall i \in \{1, \dots, d\}, |x_i| = 1\} \subseteq \text{extr}(C_\infty)$.
Let $x \in \text{extr}(C_\infty)$. Suppose that there exist i_0 such that $|x_{i_0}| < 1$. We define (y^-, y^+) such that $[y^-]_i = [y^+]_i, \forall i \neq i_0$, $[y^-]_{i_0} = -1$ and $[y^+]_{i_0} = 1$. Then x belongs to the open line segment $]y^-, y^+[$ in C_∞.

Finally, we prove that $\text{extr}(P_{\mathcal{C}}) = \{e_i\}_{i=1}^d$ the canonical basis of $\mathbb{R}^d$.
Let $i \in \{1, \ldots, d\}$ and $c \in P_{\mathcal{C}}$.

$$\langle c, e_i \rangle = c_i \leq 1$$

by definition of $P_{\mathcal{C}}$. We then have $\{e_i\}_{i=1}^d \subseteq \text{extr}(P_{\mathcal{C}})$ as c achieves this bound if and only if $c = e_i$.
$\text{extr}(P_{\mathcal{C}}) \subseteq \{e_i\}_{i=1}^d$ follows from the fact that $x = \sum_{i=1}^d x_i e_i$ is a convex combination if $|\{i \in \{1, \ldots, d\} | x_i \neq 0\}| \geq 2$.

(b) The fact that $p_2 = -\frac{a}{\|a\|_2}$ is a solution of 2 follows directly from the other side of Cauchy-Swhartz inequalities in (2).
$p_\infty = -\text{sign}(a)$ is a solution using the same decomposition as in 5.
Finally, $p_{\mathcal{P}} = e_i$, where $i = \arg\min_j a_j$ is a solution as :

$$\langle a, p \rangle \geq a_{\arg\min_j a_j} \sum_{i=1}^d p_i \tag{7}$$

$$= a_{\arg\min_j a_j} \quad \text{as } p \in C_{\mathcal{P}} \tag{8}$$

and $p = p_{\mathcal{P}}$ achieves this bound.

Question 3

An important family of functions satisfying (HEB) is the family of uniformly convex functions. Using uniform convexity and the fact that $-\nabla f(x^*) \in N_A(x^*)$., one can esasily show that (α_f, r)-uniformly convex functions satisfy a $\left(\left(\frac{r}{\alpha_f}\right)^{\frac{1}{r}}, \frac{1}{r}\right)$-(HEB):

$$f(x) - f(x^*) \geq \langle \nabla f(x^*), x - x^* \rangle + \frac{\alpha_f}{r} \|x - x^*\|^r \geq \frac{\alpha_f}{r} \|x - x^*\|^r.$$

Noticing that m-strongly convex functions are $(m, 2)$-uniformly convex, one can deduce that m-strongly convex functions satisfy a $\left(\sqrt{\frac{2}{m}}, \frac{1}{2}\right)$-(HEB)

From questions 4 to 6, we investigate the first setting discussed in the paper (Theorem 3.6) where FW with open-loop step-sizes admit acceleration rates, that is f satisfies (HEB) and the minimizer of f is in the relative interior of the feasible region C. To compare practical performance with the theory, we will consider the feasible region $C = C_{\mathcal{P}}$ and the function f defined as follow :

$$f(x) = \frac{1}{2} \|x\|_2^2$$

Question 4

In this question we prove that the point $x^* = \frac{1}{d}(1, \ldots, 1)$ is the unique minimizer of 1 using Cauchy-Schartz inequality and then show that it is in the relative interior of $C_{\mathcal{P}}$.

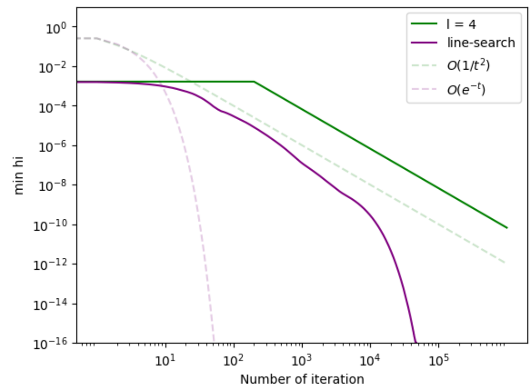
Notice that $f(x^*) = \frac{1}{2d}$. Let $c \in C_P$.

$$|\langle c, x^* \rangle| \leq \|c\| \|x^*\| \tag{9}$$

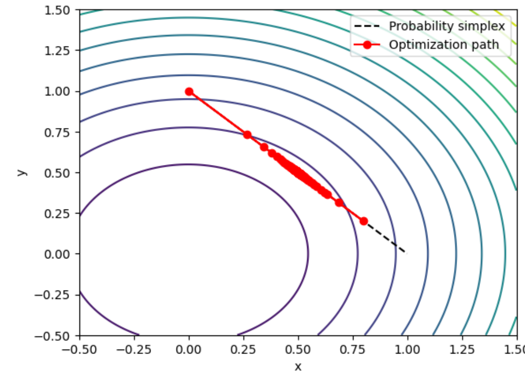
$$\frac{1}{d\|x^*\|} \leq \|c\| \quad \text{as } c \in C_P \tag{10}$$

By taking to the square, it follows that

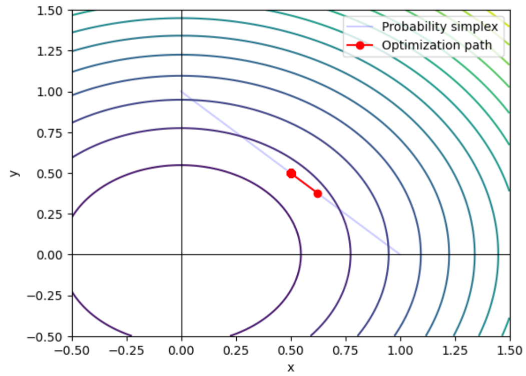
$$f(x^*) = \frac{1}{2d} \leq \frac{1}{2} \|c\|^2 = f(c) \tag{11}$$



(a) Energy-decay, $d = 100$.



(b) Optimization path, open loop $\eta_t = \frac{4}{t+4}$, $d = 2$.



(c) Optimization path, exact line-search, $d = 2$.

Figure 1: Energy-decay in log-log plot and optimization path of FW with open-loop step-size $\eta_t = \frac{4}{t+4}$ and exact line-search when the feasible region $C \subseteq \mathbb{R}^d$ is the probability simplex, the objective $f(x) = \frac{1}{2}\|x\|_2^2$ satisfy (HEB) and the minimizer of f is in the relative interior of the feasible region C. The y-axis represents the minimum primal gap. FW with open-loop step-size $\eta_t = \frac{4}{t+4}$ converges at a rate of order $O(\frac{1}{t^2})$. FW with exact line-search converges linearly.

and the inequality is strict unless there exist $\alpha \in \mathbb{R}$ such that $c = \alpha x^*$. As $c \in C_P$, α must be equal to 1 and x^* is the unique minimizer of (1).

We know prove that x^* is in the relative interior of C_P, i.e., $\mathrm{Aff}(C_P) \cap B_\beta(x^*) \subseteq C_P$ such that $\beta = \frac{1}{d}$. Let $y \in \mathrm{Aff}(C_P) \cap B_{\frac{1}{d}}(x^*)$. Suppose, for the sake of contradiction, that there exist $i_0 \in \{1, \ldots, d\}$ s.t $y_{i_0} < 0$. Then $|y_{i_0} - \frac{1}{d}| > \frac{1}{d}$ which induce $y \notin B_{\frac{1}{d}}(x^*)$:

$$\|y - x^*\| = \sqrt{\sum_{i=1}^{d}(y_i - \frac{1}{d})^2} \geq |y_{i_0} - \frac{1}{d}| > \frac{1}{d} \tag{12}$$

More ever, as $y \in \mathrm{Aff}(C_P)$, there exist $n \in \mathbb{N}^*, \lambda \in \mathbb{R}^n$, $\sum_{j=1}^{n} \lambda_j = 1$ and $x_1, \ldots, x_n \in C_P$. such that $y = \sum_{j=1}^{n} \lambda_j x_j$.

$$\begin{aligned}\sum_{i=1}^{d} y_i &= \sum_{i=1}^{d}\sum_{j=1}^{n}[x_j]_i\lambda_j \\ &= \sum_{j=1}^{n}\lambda_j\sum_{i=1}^{d}[x_j]_i \\ &= \sum_{j=1}^{n}\lambda_j \quad \text{as } x_j \in C_P \\ &= 1\end{aligned}$$

Question 5

C_P is a compact convex set of diameter $\delta = \sqrt{2}$, $f(x) = \frac{1}{2}\|x\|^2$ is a L-smooth 1-strongly convex function satysfing $\left(\mu = \sqrt{2}, \theta = \frac{1}{2}\right)$-(HEB) with unique minimizer x^* such that for $\beta = \frac{1}{d}$ $\mathrm{Aff}(C_P) \cap B_\beta(x^*) \subseteq C_P$. Let $S = \left\lceil 8L\delta^2\left(2\frac{\mu}{\beta}\right)^{1/\theta}\right\rceil = 64d^2$, $T \in \mathbb{N}$, and $\eta_t = \frac{4}{t+4}$ for all $t \in \mathbb{Z}$. Then, using Theorem 3.6, for the iterates of Frank-Wolfe algorithm with open-loop step-size with step-size η_t, it holds that

$$f(x_t) - f(x^*) \le \max\left(\left(\frac{\eta_{t-2}}{\eta_{S-1}}\right)^{\frac{1}{1-\theta}} f(x_S) - f(x^*), \left(\frac{\eta_{t-2}^2\mu L\delta^3}{\beta^2}\right)^{\frac{1}{1-\theta}} + \frac{\eta_{t-2}^2 L\delta^2}{2}\right)$$

for all $t \in \{S, \dots, T\}$. In other words, the algorithm converges at a rate of $O(\frac{1}{t^2})$.

Question 6

The results in Figure 1, show that in the setting of Theorem 3.6, FW with open-loop step-sizes $\eta_t = \frac{4}{t+4}$, converges at a rate of order $O(\frac{1}{t^2})$. This pratical result aligns with the theory.

We know consider the consider the setting when the feasible region is a polytope, the objective function is strongly convex, and the optimal solution lies in the relative interior of an at least one-dimensional face C^* of C. To compare practical performance with the theory, we will keep the feasible region $C = C_{\mathcal{P}}$ and consider the function f defined as follow :

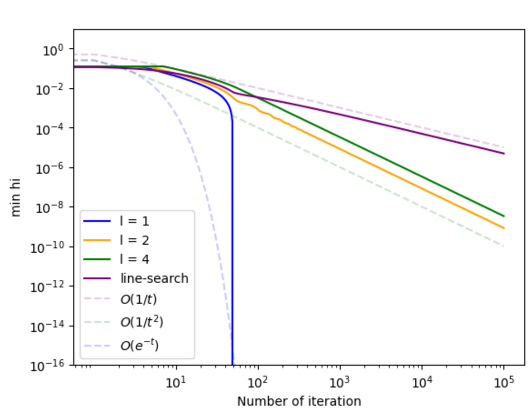
$$f(x) = \frac{1}{2}\left\|x - \rho\bar{\mathbf{1}}\right\|^2 \tag{13}$$

such that $\bar{\mathbf{1}} = (0, \dots, 0, 1, \dots, 1)$ with $\frac{d}{2}$ zeros and $\frac{d}{2}$ ones (we assume d is even), where $\rho \ge \frac{2}{d}$, and admit that the unique minimizer to 1 with f is given by

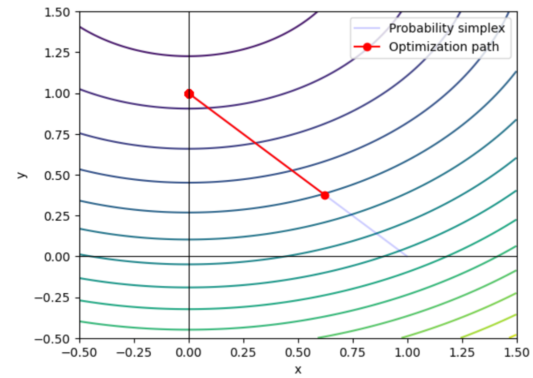
$$x^* = \frac{2}{d}\bar{\mathbf{1}}$$

Question 7

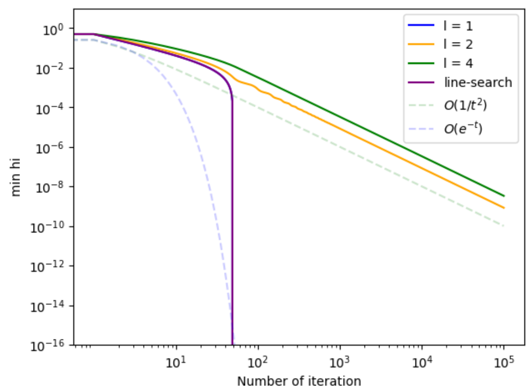
Theorem 3.6 adresses the setting where the optimal solution in the relative interior of the feasible region C, but x^* does not. To see that, one can consider the case $d = 2$ and $x^* = e_2$.
Theorem 3.10 and Theorem 3.12 adresses the setting when the feasible region C is uniformly convex, but C_P is not uniformly convex. Finally, we can neither apply Theorem 3.6, Theorem 3.10 nor Theorem 3.12.



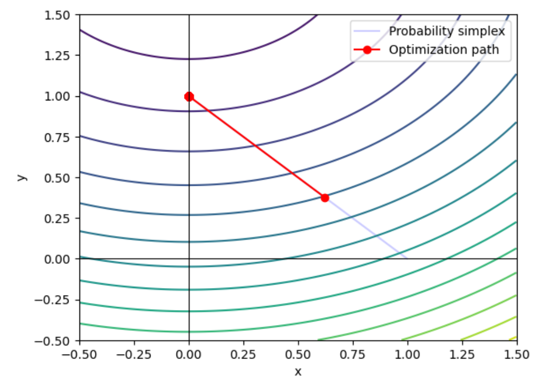
(a) Energy-decay, $d = 100$, $\rho = \frac{1}{4}$.



(b) Optimization path, open-loop $\eta_t = \frac{4}{t+4}$, $d = 2$, $\rho = \frac{1}{4}$.



(c) Energy-decay, $d = 100$, $\rho = 2$.



(d) Optimization path, line-search, $d = 2$, $\rho = 2$.

Figure 2: Energy-decay and optimization path of FW with different step-sizes when the feasible region $C \subseteq \mathbb{R}^d$ is the probability simplex, the objective $f(x) = \frac{1}{2}\|x - \bar{\mathbf{1}}\|^2$, where $\rho \in \{\frac{1}{4}, 2\}$, is strongly convex, and the optimal solution lies in the relative interior of an at least one-dimensional face of C. Figure 2a and 2c show the energy-decay when $d = 100$. The y-axis represents the minimum primal gap. Figures 2b and 2d shows the optimization path when $d = 2$. FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$ where $l \in \mathbb{N}_{\geq 2}$, converges at a rate of order $O\left(\frac{1}{t^2}\right)$. FW with open-loop step-size $\eta_t = \frac{1}{t+1}$ converges linearly. FW with line-search converges at a rate of order $O(1/t)$ when $\rho = \frac{1}{4}$ and linearly when $\rho = 2$. When convergence is linear, FW with line-search and open-loop step-size $\eta_t = \frac{1}{t+1}$ solves the problem exactly after $|\mathrm{supp}(x^*)|$ iterations.

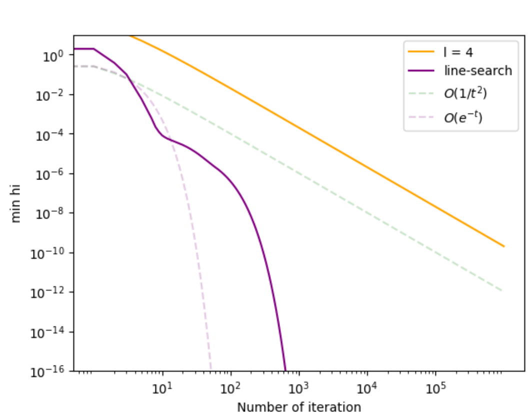
Question 8

Theorem 4.5 show that FW with open-loop step-sizes $\eta_t = \frac{4}{t+4}$ admits convergence rate of order $O\left(\frac{1}{t^2}\right)$ in the setting discussed above. As the function f defined in (13) is 1-strongly convex, the probability simplex $C_{\mathcal{P}}$ is a polytope and the unique minimizer $x^* = \frac{2}{d}\bar{\mathbf{1}}$ is contained in the relative interior of an at least one-dimensional face C^* of $C_{\mathcal{P}}$, we can apply Theorem 4.5.

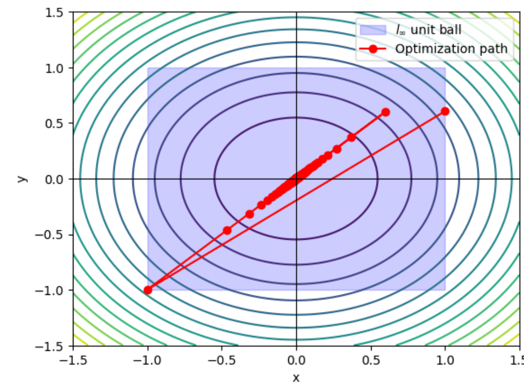
The results in Figure 2, see Section **??** for details, show that in this setting, FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$ where $l \in \mathbb{N}_{\geq 2}$, converges at a rate of order $O\left(\frac{1}{t^2}\right)$ which aligns with with the theory. FW with open-loop step-size $\eta_t = \frac{1}{t+1}$ converges linearly and solves the problem exactly after $|\mathrm{supp}(x^*)|$ iterations. The latter phenomena is not explained by the theory.

Question 9

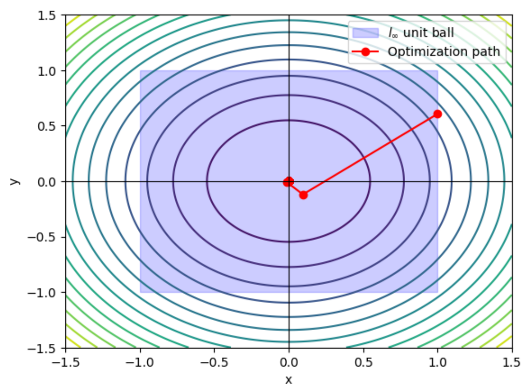
We know try to observe similar behaviors with $C = C_\infty$ as with the probability simplex. To do so, we will consider the setting explored from question 4 to 6, that is f satisfies (HEB) and the minimizer



(a) Energy-decay, $d = 100$.



(b) Optimization path, open-loop step-size $\eta_t = \frac{4}{t+4}$, $d = 2$.



(c) Optimization path, line-search, $d = 2$.

Figure 3: Energy-decay and optimization path of FW with open-loop step-size $\eta_t = \frac{4}{t+4}$ and exact line-search when the feasible region $C \subseteq \mathbb{R}^d$ is the l_∞ unit ball, the objective $f(x) = \frac{1}{2}\|x\|_2^2$ satisfy (HEB) and the minimizer of f is in the relative interior of the feasible region C. Figure 3a shows the energy-decay when d = 100. The y-axis represents the minimum primal gap. Figures 3b and 3c show the optimization path when d = 2. FW with open-loop step-size $\eta_t = \frac{4}{t+4}$ converges at a rate of order $O(\frac{1}{t^2})$. FW with exact line-search converges linearly.

of f is in the relative interior of the feasible region C. We consider the function $f(x) = \frac{1}{2}\|x\|_2^2$ whose unique minimizer is $x^* = 0$.

The results in Figure 4 show that in the setting of Theorem 3.6, FW with open-loop step-sizes $\eta_t = \frac{4}{t+4}$, converges at a rate of order $O(\frac{1}{t^2})$. This pratical result aligns with the theory and the behavior of FW on the probability simplex in this setting.

From questions 10 to 12 we consider the setting where the feasible region C is uniformly convex and the norm of the gradient of f is bounded from below by a nonnegative constant. To compare practical performance with the theory, we will consider the feasible region $C = C^2$, which is assumed to be strongly convex (or $(\alpha, 2)$ uniformly convex), and the function f defined as follow :

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2 \tag{14}$$

where $b = Ax^\dagger$ with $x^\dagger \in \text{Im}(A^T)$, $\|x^\dagger\|_2^2 = 2$ and $A \in \mathbb{R}^{N \times d}$ s.t. $N < d$.

Question 10

In this question we prove that f has lower-bounded gradient using the fact that the unconstrained minimizer of f, that is $\arg\min_{x\in\mathbb{R}^d} f(x)$, lies in the exterior of C_2, and the convexity of f.

Let's consider $A^\dagger$ the Moore-Penrose pseudo-inverse of A. We prove[1] below that $A^\dagger b$ is a minimizer of f with the smallest norm. Also $b = Ax^\dagger$, such that $x^\dagger \in \text{Im}(A^T)$, so we have $A^\dagger b = A^\dagger A x^\dagger = x^\dagger$ as $A^\dagger A$ is an orthogonal projection on $\text{Im}(A^T)$. Combining both arguments it follows that the unconstrained minimizer of f lies in the exterior of C_2 :

$$\|x\|_2 \geq \|x^\dagger\|_2 = 2 \tag{15}$$

for every $x \in \arg\min_{x\in\mathbb{R}^d} f(x)$.

Let us consider an arbitrary $x \in \mathbb{R}^d$. We write $x = z + A^\dagger b$. Then, from the properties of $A^\dagger$,

$$\begin{aligned}
f(x) &= \frac{1}{2}\|Az + AA^\dagger b - b\|_2^2 \\
&= \frac{1}{2}\|Az\|_2^2 + \langle Az, AA^\dagger b\rangle - \langle Az, b\rangle + f(A^\dagger b) \\
&= \frac{1}{2}\|Az\|_2^2 + \langle AA^\dagger Az, b\rangle - \langle Az, b\rangle + f(A^\dagger b) \\
&= \frac{1}{2}\|Az\|_2^2 + \langle Az, b\rangle - \langle Az, b\rangle + f(A^\dagger b) \\
&= \frac{1}{2}\|Az\|_2^2 + f(A^\dagger b) \\
&\geq f(A^\dagger b).
\end{aligned}$$

Therefore, $f(A^\dagger b) = \min f$. From the previous inequalities, we also see that $f(x) = f(A^\dagger b) = \min f$ if and only if $Az = 0$. In this case,

$$\begin{aligned}
\|x\|_2^2 &= \|z\|_2^2 + 2\langle z, A^\dagger b\rangle + \|A^\dagger b\|_2^2 && (16)\\
&= \|z\|_2^2 + 2\langle z, A^\dagger AA^\dagger b\rangle + \|A^\dagger b\|_2^2 && (17)\\
&= \|z\|_2^2 + 2\langle A^\dagger Az, A^\dagger b\rangle + \|A^\dagger b\|_2^2 && (18)\\
&= \|z\|_2^2 + 2\langle A^\dagger 0, A^\dagger b\rangle + \|A^\dagger b\|_2^2 && (19)\\
&= \|z\|_2^2 + \|A^\dagger b\|_2^2 \geq \|A^\dagger b\|_2^2, && (20)
\end{aligned}$$

with equality if and only if $z = 0$. Consequently, $A^\dagger b$ has minimal norm among all minimizers of f.

Now that we have proved that the unconstrained minimizer of f lies in the exterior of C^2, we use the convexity of f to deduce that f has lower-bounded gradients on C^2, that is there exist $\lambda > 0$ such that $\forall x \in C^2$

$$\|\nabla f(x)\| \geq \lambda > 0 \tag{21}$$

Let $x \in C^2$. If we suppose $\|\nabla f(x)\| = 0$, then by definition of the norm $\nabla f(x) = 0$. Hence, by convexity of f, $x \in \arg\min_{x\in\mathbb{R}^d} f(x)$ and equation (15) contradicts the fact $x \in C_2$. We then have $\|\nabla f(x)\| \geq \lambda_x > 0$ using continuity of of $\|\nabla f(x)\|$. As C_2 is closed, we deduce that :

$$\forall x \in C_2, \|\nabla f(x)\|_2 \geq \lambda = \min\{\lambda_x, x \in C_2\} > 0 \tag{22}$$

Question 11

Theorem 3.10 gives convergence rate for FW with open-loop step-sizes of order $O\left(\frac{1}{t^2}\right)$ in the setting discussed above, that is the feasible region is uniformly convex and f has lower bounded gradients. Using the last question and the assumption that C^2 is $(\alpha, 2)$ uniformly convex we can apply Theorem 3.10.

[1]The proof is take from Irene Waldspurger course note

More ever, Remark 3.11 discuss acceleration beyond rates of order $O\left(\frac{1}{t^2}\right)$ when the feasible region is strongly convex. In our setting, one can show convergence rate of order $O\left(\frac{1}{t^{l/2}}\right)$ for FW with step-sizes $\eta_t = \frac{l}{t+4}$ where $l \in \mathbb{N}_{\geq 2}$. More ever, using similar arguments, one can prove that FW with the constant open-loop step-size $\eta_t = \frac{\alpha\lambda}{2L}$ converges linearly, that is, $h_t \leq (1 - \frac{\alpha\lambda}{2L})^t h_0$ for all $t \in \{0, \ldots, T\}$.

Question 12

We know analyse numerically FW algorithms in this setting. To do so, we need to compute the optimal solution $x^* = \arg\min_{(x \in C^2} f(x)$ using KKT conditions.

We first check that the constraint is qualified using Slater's condition. The inequality constraint is convex, differentiable, and the point $0 \in C^2$ is strictly feasible as $\|0\|_2 - 1 = -1 < 0$.

We consider the Lagrangian function $\mathcal{L}$ defined as follow :

$$\forall (x, \mu) \in \mathbb{R}^d \times \mathbb{R}^+, \mathcal{L}(x, \mu) = \frac{1}{2}\|Ax - b\|_2^2 + \mu(\|x\|_2 - 1)$$

The stationnary condition $\nabla_x \mathcal{L}(x, \mu) = 0$ gives $(A^T A + \mu)x = A^T b$.

If $\mu > 0$, as $A^T A$ is positive semi-definite, the matrix $A^T A + \mu$ is invertible, and we deduce $x^* = (A^T A + \mu)^{-1} A^T b$. Also, the complementary slackness condition induce $\|x\|_2 = 1$ and we can deduce an expression to compute μ^* with a dichotomic algorithm:

$$\|(A^T A + \mu)^{-1} A^T b\|_2 - 1 = 0 \tag{23}$$

The result in Figure 4a, show that in this setting, FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$, where $l \in \mathbb{N}_{\geq 1}$, converges at a rate of order $O(1/t^l)$ which indicate a gap between theory and practice.

Question 13

In this question, we investigate the FW algorithm with linesearch:

$$\eta_t = \underset{\eta \in [0,1]}{\arg\min}\, f((1-\eta)x_t + \eta p_t) \tag{24}$$

If the objective is quadratic, e.g. $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ for $A \in \mathbb{R}^{N \times d}$, it is possible to have closed-form expressions for the linesearch.

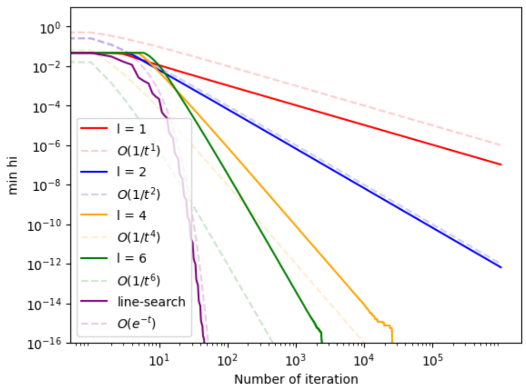
$$\begin{aligned} P(\eta) &= f((1-\eta)x_t + \eta p_t) \\ &= \frac{1}{2}\|\eta(A(p_t - x_t)) + Ax - b\|_2^2 \\ &= \eta^2 a + \eta b + c \end{aligned}$$

where $a = \|A(p_t - x_t)\|_2^2$, $b = \langle p_t - x_t, A^T A(x - x^\dagger)\rangle$ and $c = \frac{1}{2}\|Ax - b\|_2^2$.

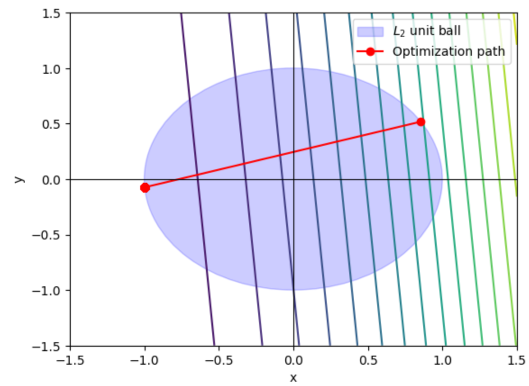
We first consider the setting of Theorem (3.6), that is f satisfies (HEB) and the minimizer of f is in the relative interior of the feasible region C. To compare practical performance with the theory, we will consider the probability simplex $C_{\mathcal{P}}$ and the l_∞ unit norm and the function $f(x) = \frac{1}{2}\|x\|_2^2$ already analysed in question 4 to 6.

The results in Figures 1a and 3a show that in the setting of Theorem 3.6, FW with exact line-search converges linearly where as FW with open-loop step-size $\eta_t = \frac{4}{t+4}$ at an order $O(1/t^2)$. This pratical result aligns with the theory. Also, when comparing the optimization paths in 1c and 3c one can see the effect of selecting the optimal step size on a line.

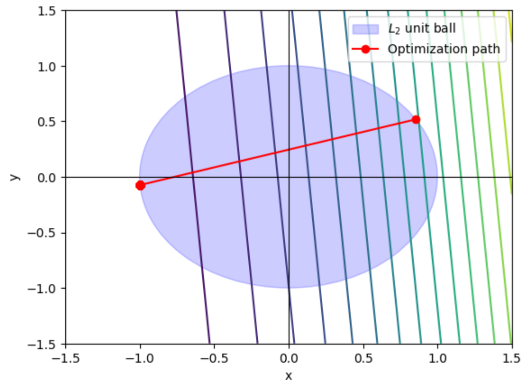
We now consider the setting of Theorem 4.5, that is the feasible region is a polytope, the objective function is strongly convex, and the optimal solution lies in the relative interior of an at least one-dimensional face C^* of C. In this setting line-search converges at a rate of order $\Omega(1/t^{1+\epsilon})$ for any



(a) Energy-decay, $d = 100$.



(b) Optimization path, open-loop $\eta_t = \frac{4}{t+4}$, $d = 2$.



(c) Optimization path, line-search, $d = 2$.

Figure 4: Energy-decay in log-log plot and optimization path of FW with different step-sizes when the feasible region $C \subseteq \mathbb{R}^d$ is the l_2 unit ball, the objective $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ is not strongly convex, and the unconstrained optimal solution $\arg\min_{x \in \mathbb{R}^d} f(x)$ lies in the exterior of C, implying that $\|\nabla f(x)\|_2 \geq \lambda > 0$ for all $x \in C$ for some $\lambda > 0$. The y-axis represents the minimum primal gap. Figure 4a shows the energy-decay when $d = 100$. Figures 4b and 4c show the optimization path when $d = 2$. FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$, where $l \in \mathbb{N}_{\geq 1}$, converges at a rate of order $O(1/t^l)$. FW with exact line-search converges linearly.

$\epsilon > 0$, and open-loop step size $\eta_t = \frac{4}{t+4}$ converges in $0t(1/t^2)$. For this experiment, we use the function $f(x) = \frac{1}{2}\left\|x - \bar{\mathbf{1}}\right\|^2$ on the probability simplex.

The results in Figure 2a and 2c show that in the setting of Theorem 4.5, FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$ where $l \in \mathbb{N}_{\geq 2}$, converges at a rate of order $O\left(\frac{1}{t^2}\right)$. FW with open-loop step-size $\eta_t = \frac{1}{t+1}$ converges linearly. FW with line-search converges at a rate of order $O(1/t)$ when $\rho = \frac{1}{4}$ and linearly when $\rho = 2$. Hence, there exist settings for which FW with open-loop step-sizes converges non-asymptotically faster than FW with line-search, which align with the theory.

We finally consider the setting of Theorem 3.10, that is the feasible region C is uniformly convex and f has lower-bounded gradient, in which FW with line-search converges linearly when the feasible region is also strongly convex, and open-loop step-size converges in $O(1/t^{l/2})$ for $l \in \mathbb{N}_{\geq 2}$. For this experience, we use $f(x) = \frac{1}{2}\|Ax - b\|_2^2$ as defined in (14), on the l_2 unit ball C_2.

The results in Figure 4a show that in the setting of Theorem 3.10, FW with open-loop step-sizes $\eta_t = \frac{l}{t+l}$, where $l \in \mathbb{N}_{\geq 1}$, converges at a rate of order $O(1/t^l)$, which is faster than the theoretical results. How ever, FW with exact line-search converges faster, i.e. linearly.

Numerical experiments details

In this section, we present the details of our numerical experiments. All of our numerical experiments are implemented in Python and performed on an Nvidia GeForce RTX 3050 Ti Laptop GPU with 4GB RAM and an Intel Core i7-11370H 8x CPU at 3.30GHz with 32 GB RAM. Our code is publicly available on GitHub. For all numerical experiments, to avoid the oscillating behavior of the primal gap, the y-axis represents $\min_{i\in\{1,\dots,t\}} h_i$, where t denotes the number of iterations and h_i the primal gap. For all energy-decay plots, we use a log-log scale. We initialize FW algorithm with x_0 a random point in the feasible region C.

references