Acceleration

Roland Becker

January 26, 2023

Contents

1 Acceleration of sequences

Algorithm 1: AGM fixed step size

Choose $x_0 \in X$, $0 \leqslant \beta, \rho \leqslant 1$. Set $x_{-1} = x_0$ and $k = 0$.

(1) $y_k = x_k + \beta(x_k - x_{k-1})$

(2) $x_{k+1} = \rho y_k$.

(3) Increment k and go to (1).

Lemma 1. *Let $\rho \leqslant 1$ and $\beta^*(\rho)$ be the solution to*

$$\frac{(1+\beta)^2}{\beta} = \frac{4}{\rho}. \tag{1}$$

Then for all $0 \leqslant \beta \leqslant \beta^*(\rho)$ *and*

$$\theta = \rho(1+\beta)\frac{1-\sqrt{1-S}}{2}, \quad q = \rho(1+\beta)\frac{1+\sqrt{1-S}}{2}, \quad S := \frac{4\rho\beta}{\rho^2(1+\beta)^2} \tag{2}$$

we have

$$x_n = \left(\theta^n + (\rho-\theta)\sum_{k=0}^{n-1} q^{n-1-k}\theta^k\right) x_0. \tag{3}$$

Proof. Let

$$z_k := x_k - \theta x_{k-1} \quad (k \geqslant 1).$$

We have

$$z_1 = (\rho-\theta)x_0, \quad z_{k+1} = q z_k \quad (k \geqslant 1)$$

if and only

$$x_{k+1} = (\theta+q)x_k - \theta q x_{k-1},$$

which leads to

$$\theta + q = \rho(1+\beta), \quad \theta q = \rho\beta$$

Taking squares and subtracting we have

$$(\theta-q)^2 = \rho^2(1+\beta)^2 - 4\rho\beta \quad \left(= \rho^2(1-\beta)^2 - 4\rho\beta(1-\rho)\right).$$

The function $\phi(x) = (1+x)^2/x$ is strictly decreasing and convex on $]0,1]$ and $\phi(1) = 4$. So for β satisfying (1) we have $S \leqslant 1$.

$$2\theta = \rho(1+\beta) - \sqrt{\rho^2(1+\beta)^2 - 4\rho\beta} \;\Rightarrow\; (2)$$

This implies with $\beta \leqslant 1$ that $\theta \leqslant \rho$. Finally, it is clear that (3) holds for $n = 0$. Then by induction

$$\begin{aligned} x_{n+1} =& z_{n+1} + \theta x_n = q^n z_1 + \theta x_n = q^n(\rho-\theta)x_0 + \left(\theta^{n+1} + (\rho-\theta)\sum_{k=0}^{n-1} q^{n-1-k}\theta^{k+1}\right) x_0 \\ &= \left(\theta^{n+1} + (\rho-\theta)\sum_{k=1}^{n} q^{n-k}\theta^k + q^n(\rho-\theta)\right) x_0 = \left(\theta^{n+1} + (\rho-\theta)\sum_{k=0}^{n} q^{n-k}\theta^k\right) x_0 \end{aligned}$$

□

It follows from (3) that

$$x_n = \left(\theta^n + (\rho - \theta)\sum_{k=0}^{n-1}\left(\frac{\theta}{q}\right)^k q^{n-1}\right) x_0.$$

This expression is minimized if β is chosen such that $q = \theta$.

Theorem 1. *Let $\beta = \beta^*$. Then we have*

$$\beta = \frac{2 - \rho - 2\sqrt{1-\rho}}{\rho} = 1 + \frac{2\sqrt{1-\rho}\left(\sqrt{1+\rho} - 1\right)}{\rho} \tag{4}$$

$$x_n = \left(\frac{\rho}{1 + \sqrt{1-\rho}}\right)^n \left(1 + 2n\sqrt{1-\rho}\right) x_0. \tag{5}$$

Remark 1. *For $\rho = 1 - 1/\kappa$, we find $\rho^* = 1 - \frac{1+\kappa^{-\frac{1}{2}}}{1+\kappa^{\frac{1}{2}}} \approx 1 - 1/\sqrt{\kappa}$ and $\beta^* \approx 1 - 2/\sqrt{\kappa}$.*

Proof. From (3) we get

$$x_n = \theta^n \left(1 + n\frac{(\rho - \theta)}{\theta}\right) x_0.$$

We have

$$\theta = \rho\frac{1+\beta}{2}, \quad \frac{(\rho-\theta)}{\theta} = \frac{1-\beta}{1+\beta}$$

and

$$\frac{(1+\beta)^2}{\beta} = \frac{4}{\rho}, \quad \frac{(1-\beta)^2}{\beta} = \frac{4(1-\rho)}{\rho} \quad \Rightarrow \quad \frac{1-\beta}{1+\beta} = 2\sqrt{1-\rho}$$

and

$$\beta^2 + (2 - \frac{4}{\rho})\beta = -1 \quad \Rightarrow \quad \left(\beta - \frac{2-\rho}{\rho}\right)^2 = \frac{(2-\rho)^2 - \rho^2}{\rho^2} = \frac{4(1-\rho)}{\rho^2}$$

so we get (4). We have $\beta \geqslant 0$ since $\sqrt{1-x} \leqslant 1 - x/2$ and $\beta \leqslant 1$ since $1 - \rho \leqslant \sqrt{1-\rho}$. We finally have

$$\theta = \rho\frac{1 + \frac{2-\rho}{\rho} - \frac{2\sqrt{1-\rho}}{\rho}}{2} = \frac{\rho + 2 - \rho - 2\sqrt{1-\rho}}{2} = 1 - \sqrt{1-\rho}$$

□

2 Accelerated gradient methods

2.1 Constant step size

Algorithm 2: AGM fixed step size

Choose $x_0 \in X$, $0 \leqslant \beta \leqslant 1$. Set $x_{-1} = x_0$ and $k = 0$.

(1) $y_k = x_k + \beta(x_k - x_{k-1})$

(2) $x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k)$.

(3) Increment k and go to (1).

Let us start with, for any $x \in X$,

$$\begin{cases} f(x_{k+1}) \leqslant & f(y_k) - \dfrac{1}{2L}\|\nabla f(y_k)\|^2 \\ f(x) \geqslant & f(y_k) + \langle \nabla f(y_k), x - y_k\rangle + \dfrac{\mu}{2}\|x - y_k\|^2 \end{cases}$$

$$\Rightarrow \quad f(x_{k+1}) - f(x) \leqslant \langle \nabla f(y_k), y_k - x\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2 - \frac{\mu}{2}\|x - y_k\|^2$$

Let

$$u_k := \frac{y_k - (1-\theta)x_k}{\theta} = x_k + \frac{y_k - x_k}{\theta} = x_k + \frac{\beta}{\theta}(x_k - x_{k-1})$$

$$v_k := \frac{x_k - (1-\theta)x_{k-1}}{\theta} = x_k + \frac{(1-\theta)(x_k - x_{k-1})}{\theta} = x_k + \frac{(1-\theta)(y_k - x_k)}{\theta\beta}$$

Using $2ab - a^2 = b^2 - (a-b)^2$ we have

$$\langle \nabla f(y_k), y_k - (1-\theta)x_k - \theta x^*\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2 = \frac{L\theta^2}{2}\left(\|u_k - x^*\|^2 - \|v_{k+1} - x^*\|^2\right)$$

and

$$\begin{aligned} \theta\|y_k - x^*\|^2 + (1-\theta)\|y_k - x_k\|^2 = & \|y_k - (1-\theta)x_k - \theta x^*\|^2 + \theta(1-\theta)\|x_k - x^*\|^2 \\ = & \theta^2\|u_k - x^*\|^2 + \theta(1-\theta)\|x_k - x^*\|^2 \end{aligned}$$

Then with $\rho := 1 - 1/\kappa_f$

$$\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant \frac{\theta^2 L}{2}\left(\rho\|u_k - x^*\|^2 - \|v_{k+1} - x^*\|^2\right) - \frac{\theta(1-\theta)\mu}{2}\|x_k - x^*\|^2$$

Now let

$$\beta \leqslant \lambda \leqslant 1 - \theta, \quad w_k = x_k + \frac{\lambda}{\theta}(x_k - x_{k-1})$$

Then

$$\|u_k - x^*\|^2 = \|w_k - x^*\|^2 + 2\langle w_k - x^*, u_k - w_k\rangle + \|u_k - w_k\|^2$$

since $u_k - w_k = \frac{\beta-\lambda}{\theta}(x_k - x_{k-1}) = \frac{\beta-\lambda}{\theta}\frac{\theta}{\lambda}(w_k - x_k)$

$$\begin{aligned} 2\langle w_k - x^*, u_k - w_k\rangle =& 2\frac{\beta-\lambda}{\lambda}\langle w_k - x^*, w_k - x_k\rangle \\ =& \frac{\beta-\lambda}{\lambda}\left(\|w_k - x^*\|^2 + \|w_k - x_k\|^2 - \|x_k - x^*\|^2\right) \end{aligned}$$

so

$$\begin{aligned} \|u_k - x^*\|^2 &= \|w_k - x^*\|^2 + 2\langle w_k - x^*, u_k - w_k\rangle + \|u_k - w_k\|^2 \\ &= \|w_k - x^*\|^2 + \frac{\beta-\lambda}{\lambda}\left(\|w_k - x^*\|^2 + \|w_k - x_k\|^2 - \|x_k - x^*\|^2\right) + \|u_k - w_k\|^2 \\ &= \left(1 - \frac{\lambda-\beta}{\lambda}\right)\|w_k - x^*\|^2 + \frac{\lambda-\beta}{\lambda}\|x_k - x^*\|^2 - \frac{\beta(\lambda-\beta)}{\theta^2}\|x_k - x_{k-1}\|^2 \end{aligned}$$

Similarly with $v_k - w_k = \frac{1-\theta-\lambda}{\theta}(x_k - x_{k-1}) = \frac{1-\theta-\lambda}{\theta}\frac{\theta}{\lambda}(w_k - x_k)$

$$\begin{aligned} 2\langle w_k - x^*, v_k - w_k\rangle =& 2\frac{1-\theta-\lambda}{\lambda}\langle w_k - x^*, w_k - x_k\rangle \\ =& \frac{1-\theta-\lambda}{\lambda}\left(\|w_k - x^*\|^2 + \|w_k - x_k\|^2 - \|x_k - x^*\|^2\right) \end{aligned}$$

$$\begin{aligned} \|v_k - x^*\|^2 &= \|w_k - x^*\|^2 + 2\langle w_k - x^*, v_k - w_k\rangle + \|v_k - w_k\|^2 \\ &= \|w_k - x^*\|^2 + \frac{1-\theta-\lambda}{\lambda}\left(\|w_k - x^*\|^2 + \|w_k - x_k\|^2 - \|x_k - x^*\|^2\right) + \|v_k - w_k\|^2 \\ &= \left(1 + \frac{1-\theta-\lambda}{\lambda}\right)\|w_k - x^*\|^2 - \frac{1-\theta-\lambda}{\lambda}\|x_k - x^*\|^2 + \frac{(1-\theta)(1-\theta-\lambda)}{\theta^2}\|x_k - x_k \end{aligned}$$

Then we have

$$\begin{aligned} \rho\|u_k - x^*\|^2 - \|v_{k+1} - x^*\|^2 = \rho\left(1 - \frac{\lambda-\beta}{\lambda}\right)\|w_k - x^*\|^2 - \left(1 + \frac{1-\theta-\lambda}{\lambda}\right)\|w_{k+1} - x^*\|^2 \\ +\rho\frac{\lambda-\beta}{\lambda}\|x_k - x^*\|^2 + \frac{1-\theta-\lambda}{\lambda}\|x_{k+1} - x^*\|^2 \\ -\rho\frac{\beta(\lambda-\beta)}{\theta^2}\|x_k - x_{k-1}\|^2 - \frac{(1-\theta)(1-\theta-\lambda)}{\theta^2}\|x_{k+1} - x_k\|^2 \end{aligned}$$

TEST: for $\lambda = 1-\theta$ we have $w_k = v_k$ and

$$\rho\left\|u_k - x^*\right\|^2 - \left\|v_{k+1} - x^*\right\|^2 = \rho\left(1 - \frac{\lambda-\beta}{\lambda}\right)\left\|v_k - x^*\right\|^2 - \left\|v_{k+1} - x^*\right\|^2 + \rho\frac{\lambda-\beta}{\lambda}\left\|x_k - x^*\right\|^2 - \rho\frac{\beta(\lambda-\beta)}{\theta^2}\left\|x_k - x_{k-1}\right\|^2$$

.... ALL IN Xk ....

We have

$$\begin{aligned}
\left\|u_k - x^*\right\|^2 &= \left\|x_k - x^*\right\|^2 + 2\langle x_k - x^*, u_k - x_k\rangle + \left\|u_k - x_k\right\|^2 \\
&= \left\|x_k - x^*\right\|^2 + 2\frac{\beta}{\theta}\langle x_k - x^*, x_k - x_{k-1}\rangle + \frac{\beta^2}{\theta^2}\left\|x_k - x_{k-1}\right\|^2 \\
&= \left\|x_k - x^*\right\|^2 + \frac{\beta}{\theta}\left(\left\|x_k - x^*\right\|^2 + \left\|x_k - x_{k-1}\right\|^2 - \left\|x_{k-1} - x^*\right\|^2\right) + \frac{\beta^2}{\theta^2}\left\|x_k - x_{k-1}\right\|^2 \\
&= \frac{\beta+\theta}{\theta}\left\|x_k - x^*\right\|^2 + \frac{\beta(\beta+\theta)}{\theta^2}\left\|x_k - x_{k-1}\right\|^2 - \frac{\beta}{\theta}\left\|x_{k-1} - x^*\right\|^2
\end{aligned}$$

and similarly

$$\left\|v_k - x^*\right\|^2 = \frac{1}{\theta}\left\|x_k - x^*\right\|^2 + \frac{1-\theta}{\theta^2}\left\|x_k - x_{k-1}\right\|^2 - \frac{1-\theta}{\theta}\left\|x_{k-1} - x^*\right\|^2$$

such that

$$\begin{aligned}
\rho\left\|u_k - x^*\right\|^2 - \left\|v_{k+1} - x^*\right\|^2 =& \frac{1-\theta+\rho(\beta+\theta)}{\theta}\left\|x_k - x^*\right\|^2 - \frac{1}{\theta}\left\|x_{k+1} - x^*\right\|^2 - \frac{\beta\rho}{\theta}\left\|x_{k-1} - x^*\right\|^2 \\
&+ \frac{\rho\beta(\beta+\theta)}{\theta^2}\left\|x_k - x_{k-1}\right\|^2 - \frac{1-\theta}{\theta^2}\left\|x_{k+1} - x_k\right\|^2
\end{aligned}$$

Then with

$$\begin{aligned}
\frac{\theta L}{2}\left(1-\theta+\rho(\beta+\theta)\right) - \frac{\mu\theta(1-\theta)}{2} =& \frac{\theta L}{2}\left(1-\theta+\rho(\beta+\theta) - \frac{1-\theta}{\kappa_f}\right) \\
=& \frac{\theta L}{2}\left(\rho(1-\theta)+\rho(\beta+\theta)\right) = \frac{\theta L}{2}\left(\rho(1+\beta)\right)
\end{aligned}$$

and $e_k := \frac{L}{2}\left\|x_k - x^*\right\|^2$, $d_k := \frac{\theta L}{2}\left\|x_k - x^*\right\|^2$ we have

$$\Delta f_{k+1} \leqslant (1-\theta)\Delta f_k + \theta\left(\rho(1+\beta)e_k - e_{k+1} - \beta\rho e_{k-1}\right) + \left(\rho\beta(\beta+\theta)d_k - (1-\theta)d_{k+1}\right)$$

or with $\alpha := \rho\beta(\beta+\theta)$

$$\begin{aligned}
\theta\Delta f_{k+1} + (1-\rho)e_{k+1} + (1-\theta-\alpha)\frac{L}{2}\left\|x_{k+1} - x_k\right\|^2 \leqslant& (1-\theta)\left(\Delta f_k - \Delta f_{k+1}\right) \\
&+ \rho(e_k - e_{k+1}) - \beta\rho(e_{k-1} - e_k) + \frac{L\alpha}{2}\left(\left\|x_k - x_{k-1}\right\|^2 - \left\|x_{k+1} - x_k\right\|^2\right)
\end{aligned}$$

Let $a_k := \theta\Delta f_k + (1-\rho)e_k + (1-\theta-\alpha)\frac{L}{2}\left\|x_k - x_{k-1}\right\|^2$. Then

$$\begin{aligned}\sum_{k=n+1}^{\infty} a_k \leqslant &(1-\theta)\Delta f_n + \rho e_n - \beta\rho e_{n-1} + \frac{L\alpha}{2}\left\|x_n - x_{n-1}\right\|^2 \\ &\leqslant \max\left\{\frac{1-\theta}{\theta}, \frac{\rho}{1-\rho}, \frac{\alpha}{1-\theta-\alpha}\right\} a_n\end{aligned}$$

. . . . WORKS . . . .

Let

$$u_k := \frac{1}{\theta}\left(y_k - (1-\theta)x_k\right) = x_k + \frac{y_k - x_k}{\theta}$$

Using $2ab - a^2 = b^2 - (a-b)^2$ we have

$$\langle\nabla f(y_k), \theta u_k - \theta x^*\rangle - \frac{1}{2L}\left\|\nabla f(y_k)\right\|^2 = \frac{L}{2}\left(\left\|\theta u_k - \theta x^*\right\|^2 - \left\|x_{k+1} - (1-\theta)x_k - \theta x^*\right\|^2\right)$$

Let

$$v_k := \frac{x_k}{\theta} - \frac{(1-\theta)x_{k-1}}{\theta} = x_k + \frac{(1-\theta)(x_k - x_{k-1})}{\theta} = x_k + \frac{(1-\theta)(y_k - x_k)}{\theta\beta}$$

and $\Delta f_k := f(x_k) - f^*$. We then have with $0 < \theta < 1$

$$\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant \frac{L\theta^2}{2}\left(\left\|u_k - x^*\right\|^2 - \left\|v_{k+1} - x^*\right\|^2\right) - \frac{\theta\mu}{2}\left\|x^* - y_k\right\|^2 - \frac{(1-\theta)\mu}{2}\left\|x_k - y_k\right\|^2$$

Next we have

$$\begin{aligned} v_k &= y_k + \frac{(1-\theta-\theta\beta)(y_k - x_k)}{\theta\beta}, \quad u_k = y_k + \frac{(1-\theta)(y_k - x_k)}{\theta} \\ v_k - u_k &= \frac{(1-\theta-\beta)(y_k - x_k)}{\theta\beta} = \lambda(v_k - y_k), \quad \lambda := \frac{1-\theta-\beta}{1-\theta-\theta\beta},\end{aligned}$$

such that

$$\begin{aligned}\left\|u_k - x^*\right\|^2 &= \left\|v_k - x^*\right\|^2 - 2\langle v_k - x^*, v_k - u_k\rangle^2 + \left\|v_k - u_k\right\|^2 \\ &= \left\|v_k - x^*\right\|^2 - 2\lambda\langle v_k - x^*, v_k - y_k\rangle^2 + \left\|v_k - u_k\right\|^2 \\ &= \left\|v_k - x^*\right\|^2 - \lambda\left(\left\|v_k - x^*\right\|^2 + \left\|v_k - y_k\right\|^2 - \left\|y_k - x^*\right\|^2\right) + \left\|v_k - u_k\right\|^2 \\ &= (1-\lambda)\left\|v_k - x^*\right\|^2 - \lambda(1-\lambda)\left\|v_k - y_k\right\|^2 + \lambda\left\|y_k - x^*\right\|^2\end{aligned}$$

It follows that

$$\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant \frac{L\theta^2}{2}\left((1-\lambda)\left\|\nu_k - x^*\right\|^2 - \left\|\nu_{k+1} - x^*\right\|^2\right)$$
$$+\left(\frac{L\theta^2}{2}\lambda - \frac{\theta\mu}{2}\right)\left\|x^* - y_k\right\|^2 - \left(\frac{(1-\theta)\mu}{2} + \frac{L(1-\theta-\beta)(1-\theta)}{2\beta}\right)\left\|x_k - y_k\right\|^2$$

since

$$\lambda(1-\lambda)\left(\frac{1-\theta-\theta\beta}{\theta\beta}\right)^2 = \frac{1-\theta-\beta}{1-\theta-\theta\beta}\frac{\beta(1-\theta)}{1-\theta-\theta\beta}\left(\frac{1-\theta-\theta\beta}{\theta\beta}\right)^2 = \frac{(1-\theta-\beta)(1-\theta)}{\theta^2\beta}$$

We now chose θ such that $\lambda = \theta$, i.e.

$$\theta^{-1} = \frac{1-\theta-\theta\beta}{1-\theta-\beta} = 1 + \frac{\beta(1-\theta)}{1-\theta-\beta} \quad \Rightarrow \quad \frac{1-\theta}{\theta} = \frac{\beta(1-\theta)}{1-\theta-\beta} \quad \Rightarrow$$
$$1-\theta-\beta = \theta\beta \quad \Rightarrow \quad \theta = \frac{1-\beta}{1+\beta} \quad \Rightarrow \quad \beta = \frac{1-\theta}{1+\theta}$$

Then

$$\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant \frac{L\theta^2}{2}\left((1-\theta)\left\|\nu_k - x^*\right\|^2 - \left\|\nu_{k+1} - x^*\right\|^2\right)$$
$$+\left(\frac{L\theta^3}{2} - \frac{\theta\mu}{2}\right)\left\|x^* - y_k\right\|^2 - \left(\frac{(1-\theta)\mu}{2} + \frac{L\theta(1-\theta)}{2}\right)\left\|x_k - y_k\right\|^2$$

Proposition 1. *Suppose that* f *is* μ*-strongly convex and* ∇f *is* L*-Lipschitz and let* $\kappa_f := L/\mu$*. Set*

$$\theta := \frac{1-\beta}{1+\beta}, \quad \nu_k = x_k + \frac{y_k - x_k}{\theta}, \quad e_k := \Delta f_k + \frac{L\theta^2}{2}\left\|\nu_k - x^*\right\|^2 \tag{6}$$

Under the condition

$$\theta \leqslant \kappa_f^{-\frac{1}{2}} \tag{7}$$

we have

$$e_{k+1} \leqslant (1-\theta)e_k - \frac{1-\theta}{2}(\mu + L\theta)\left\|x_k - y_k\right\|^2 \tag{8}$$

In case $\mu = 0$ we cannot satisfy (7). But we have

$$\begin{aligned}\left\|x^* - y_k\right\|^2 &= \left\|x^* - x_k\right\|^2 + 2\beta\langle x^* - x_k, x_k - x_{k-1}\rangle + \left\|y_k - x_k\right\|^2 \\ &= \left\|x^* - x_k\right\|^2 + \beta\left(\left\|x^* - x_{k-1}\right\|^2 - \left\|x^* - x_k\right\|^2 - \left\|x_k - x_{k-1}\right\|^2\right) + \left\|y_k - x_k\right\|^2\end{aligned}$$

2.2 Acceleration of sequences

Algorithm 3: Acceleration fixed step

Inputs: $x_0 \in X$, $0 < \rho < 1$, $0 \leqslant \beta \leqslant 1$, Set $k = 0$.

(1) $x_{k+1} = \rho((1+\beta)x_k - \beta x_{k-1})$.

(4) Increment k and go to (1).

Classical two-step analysis

$$\begin{bmatrix} x_{k+1} \\ x_k \end{bmatrix} = \rho \begin{bmatrix} 1+\beta & -\beta \\ \rho^{-1} & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ -(1+\beta) & 1 \end{bmatrix} \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & \rho \\ -\beta & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ -(1+\beta) & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & \rho \\ -\beta & 0 \end{bmatrix} = \begin{bmatrix} 0 & \rho \\ -\beta & \rho(1+\beta) \end{bmatrix}$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & \rho \\ -\beta & \rho(1+\beta) \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix} \qquad y_{k+1} = (1+\beta)x_{k+1} - \beta x_k$$

$$\lambda^2 - \rho(1+\beta)\lambda = -\beta\rho \quad \Leftrightarrow \quad \left(\lambda - \frac{\rho(1+\beta)}{2}\right)^2 = \frac{\rho^2(1+\beta)^2}{4} - \frac{4\beta\rho}{4} = \frac{\rho^2 + \beta^2 + 2\beta\rho(\rho-2)}{4}$$

If

$$\rho(1+\beta)^2 = 4\beta \quad \Leftrightarrow \quad \beta^2 - 2\beta(2/\rho - 1) = -1 \quad \Leftrightarrow \quad (\beta - (2/\rho - 1))^2 = (2/\rho - 1)^2 - 1$$

we have

$$\lambda = \frac{\rho(1+\beta)}{2} = \frac{2\beta}{1+\beta} \tag{9}$$

$$\lambda^2 - (1+\beta)\lambda = -\rho^{-1}\beta$$

$$\left(\lambda - \frac{1+\beta}{2}\right)^2 = \frac{1 + 2\beta + \beta^2 - 4\rho^{-1}\beta}{4}$$

$$1 + 2\beta + \beta^2 - 4\rho^{-1}\beta \geqslant 0 \quad \Leftrightarrow \quad \left(\beta + (1 - 2\rho^{-1})\right)^2 \geqslant (1 - 2\rho^{-1})^2 - 1 = 4(\rho^{-2} - \rho^{-1})$$
$$\Leftrightarrow \quad \beta \geqslant (2\rho^{-1} - 1) \pm 2\rho^{-1}\sqrt{1-\rho}$$

$$1 + 2\beta + \beta^2 - 4\rho^{-1}\beta \geqslant 0 \quad \Leftrightarrow \quad (1-\beta)^2 \geqslant 4(\rho^{-1} - 1)\beta$$

$$\phi(x) = \frac{(1-x)^2}{x} \quad (0 < x < 1) \quad y = \phi(x) \quad \Leftrightarrow \quad -1 = -2x - xy + x^2$$
$$\Leftrightarrow \quad (x - (1 + y/2))^2 = (1 + y/2)^2 - 1 = y^2/4 + y$$
$$x = 1 + y/2 - \sqrt{y^2/4 + y} = 1 - \frac{\sqrt{y^2 + 4y} - y}{2}$$
$$\phi'(x) = \frac{x^2 - 1}{x^2}, \quad \phi''(x) = 2x^{-3}$$

Then

$$\lambda = \frac{1+\beta}{2} \pm \frac{\sqrt{1 + 2\beta + \beta^2 - 4\rho^{-1}\beta}}{2} = \frac{1+\beta}{2} \pm \frac{\sqrt{(1-\beta)^2 - 4(\rho^{-1} - 1)\beta}}{2}$$

Suppose this minimized, if the square root is zero.

$$\beta = 1 - \frac{\sqrt{y^2 + 4y} - y}{2}, \quad y = 4(\rho^{-1} - 1) = 4\frac{1-\rho}{\rho},$$
$$y^2 + 4y = \frac{16(1-\rho)^2 + 16\rho(1-\rho)}{\rho^2} = 16(1-\rho)\frac{1-\rho+\rho}{\rho^2} = \frac{16(1-\rho)}{\rho^2}$$
$$\beta = 1 - \frac{1}{2}\left(\frac{4\sqrt{1-\rho}}{\rho} - 4\frac{1-\rho}{\rho}\right) = 1 - 2\frac{\sqrt{1-\rho} - (1-\rho)}{\rho}$$

and

$$\lambda = \frac{1+\beta}{2} = 1 - \frac{\sqrt{1-\rho} - (1-\rho)}{\rho}$$

If $\rho = 1 - 1/K = (K-1)/K$

$$\beta = 1 - 2\frac{\sqrt{1-\rho} - (1-\rho)}{\rho} = 1 - 2(K^{-\frac{1}{2}} - K^{-1})\frac{K}{K-1} = 1 - 2\frac{K^{\frac{1}{2}} - 1}{K-1} = 1 - \frac{2}{K^{\frac{1}{2}} + 1}$$
$$\lambda = 1 - \frac{1}{K^{\frac{1}{2}} + 1}$$

Eigenvector

$$\begin{bmatrix} \rho\lambda \\ 1 \end{bmatrix}$$

2.3 Accelerated gradient method

We will use the following fact about geometrical convergence.

Lemma 2. *Let* $a_n \geqslant 0$*,* $n \in \mathbb{N}$*. Then under the condition that there is* $C \geqslant 0$ *such that*

$$\sum_{k=n+1}^{\infty} a_k \leqslant C a_n \quad \forall n \in \mathbb{N} \tag{10}$$

we have

$$a_{m+n} \leqslant (C+1)\rho^m a_n \quad \forall m, n \in \mathbb{N}, \quad \rho = \frac{C}{C+1}. \tag{11}$$

Proof. Let $S_n := \sum_{k=n}^{\infty} a_k$. By (10) we have

$$S_{n+1} \leqslant C\left(S_n - S_{n+1}\right) \quad \Rightarrow \quad S_{n+1} \leqslant \rho S_n.$$

Then it follows again from (10) by induction that

$$S_{n+m} \leqslant \rho^m S_n \quad \Rightarrow \quad a_{n+m} \leqslant S_{n+m} \leqslant \rho^m S_n = \rho^m(a_n + S_{n+1}) \leqslant (C+1)\rho^m a_n.$$

□

We will use the following generalization of Lemma 2.

Lemma 3. *Let* $a_n \geqslant 0$*,* $n \in \mathbb{N}$*. Under the condition that there is* $C > 0, D \geqslant 0$ *such that*

$$\sum_{k=1}^{\infty} a_k \leqslant C a_0, \quad \sum_{k=n+1}^{\infty} a_k \leqslant C a_n + D a_{n-1} \quad \forall n \in \mathbb{N}_1 \tag{12}$$

there exists $0 \leqslant \beta < 1$ *such that*

$$(1-\beta)a_{n+1} + \beta a_n \leqslant \rho^n (C+\beta) a_0 \tag{13}$$

with

$$\rho = 1 - 1/E, \quad E \leqslant C + D + \frac{D}{C+D}, \quad \beta = D/E. \tag{14}$$

Proof. Let $S_n := \sum_{k=n}^{\infty} a_k$. By (12) we have $S_1 \leqslant C(S_0 - S_1)$ and for $n \geqslant 1$

$$S_{n+1} \leqslant C\left(S_n - S_{n+1}\right) + D\left(S_{n-1} - S_n\right) \quad \Rightarrow \quad (C+1)S_{n+1} \leqslant (C-D)S_n + DS_{n-1}$$

Let for $n \geqslant 1$ and $\beta \in \mathbb{R}$

$$\widetilde{S}_n := (1-\beta)S_n + \beta S_{n-1}$$

Then we wish to find $E \geqslant 0$ such that for $n \geqslant 1$

$$(E+1)\widetilde{S}_{n+1} \leqslant E\widetilde{S}_n \tag{15}$$

which amounts to

$$(1-\beta)(E+1)S_{n+1} \leqslant (E-\beta-2\beta E)S_n + \beta E S_{n-1}$$

Choosing

$$E = \sqrt{D + \frac{(C+D)^2}{4}} + \frac{C+D}{2}, \quad \beta = \sqrt{D + \frac{(C+D)^2}{4}} - \frac{C+D}{2} \tag{16}$$

we have $\beta E = D$ and $E - C - D = \beta$, which shows (15). We have with $\sqrt{1+2x} \leqslant 1 + x$

$$E = \frac{C+D}{2}\left(1 + \sqrt{1 + \frac{4D}{(C+D)^2}}\right) \leqslant \frac{C+D}{2}\left(2 + \frac{2D}{(C+D)^2}\right) = C + D + \frac{D}{C+D}.$$

Since $E > D$ we have $\beta < 1$. From (15) we find for $n \geqslant 1$

$$\widetilde{S}_{n+1} \leqslant \rho^n \widetilde{S}_1$$

Then

$$\begin{aligned}(1-\beta)a_{n+1} + \beta a_n \leqslant & \widetilde{S}_n \leqslant \rho^n \widetilde{S}_1 = \rho^n\left((1-\beta)S_1 + \beta S_0\right) \\ = & \rho^n\left((1-\beta)S_1 + \beta(S_1 + a_0)\right) = \rho^n (C+\beta)a_0\end{aligned}$$

□

Algorithm 4: AGM with line search

Inputs: $x_0 \in X$, $t_0 > 0$. Set $y_0 = x_0$ $k = 0$.

(1) While $f(x_Q^*(t_k, y_k) > Q^*(t_k, y_k) : t_k = t_k/2$.

(2) $x_{k+1} = x_Q^*(t_k, y_k)$.

(3) $y_{k+1} = x_{k+1} + \beta_k(x_{k+1} - x_k)$.

(4) $t_{k+1} = 2 * t_k$.

(5) Increment k and go to (1).

We have

$$y_{k+1} = y_k - t_k \nabla f(y_k) + \beta_k(x_{k+1} - x_k)$$

We have

$$\begin{aligned}
f(x_{k+1}) - f(x_k) \leqslant& \frac{1}{2t_k}\left(\|y_k - x_k\|^2 - \|x_{k+1} - x_k\|^2\right)\\
=& \frac{1}{2t_k}\left(\beta^2\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x_k\|^2\right)\\
\leqslant& \frac{\beta^2}{2L}\left(\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x_k\|^2\right) - \frac{1-\beta^2}{2L}\|x_{k+1} - x_k\|^2
\end{aligned}$$

and

$$\begin{aligned}
\frac{\mu}{2}\|x_{k+1} - x^*\|^2 \leqslant& f(x_{k+1}) - f(x^*) \leqslant \frac{1}{2t_k}\left(\|y_k - x^*\|^2 - \|x_{k+1} - x^*\|\right)\\
\leqslant& \frac{L}{2}\left((1+\delta)\|x_k - x^*\|^2 - \|x_{k+1} - x^*\| + (1+\delta^{-1})\beta_{k-1}^2\|x_{k+1} - x_k\|^2\right)
\end{aligned}$$

so

$$\|x_{k+1} - x^*\|^2 \leqslant \frac{\kappa}{\kappa+1}\left((1+\delta)\|x_k - x^*\|^2 + (1+\delta^{-1})\beta_{k-1}^2\|x_{k+1} - x_k\|^2\right)$$

and

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 + \frac{(1+\delta^{-1})\beta 2L}{1-\beta^2}\left(f(x_{k+1}) - f(x_k)\right) \leqslant& \frac{(1+\delta^{-1})\beta^3}{1-\beta^2}\left(\|x_k - x_{k-1}\|^2 - \|x_{k+1} - x_k\|^2\right)\\
&+ (1+\delta)\frac{\kappa}{\kappa+1}\|x_k - x^*\|^2
\end{aligned}$$

2.4 AGM for quadratyics

Let

$$f(x) = \frac{1}{2}x^T A x \tag{17}$$

and consider

$$\begin{cases} x_{k+1} = y_k - tAy_k \\ y_{k+1} = y_k + \beta(y_k - x_k) - sAy_k \end{cases} \tag{18}$$

The iteration reads in matrix-form

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & I - tA \\ -\beta & (1+\beta)I - sA \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

We have

$$y_{k+1} = (1+\beta)(I - tA)y_k - \beta(I - tA)y_{k-1}$$

Les racines de

$$y^2 - (1+\beta)\theta y = -\beta\theta$$

sont

$$y = \frac{(1+\beta)\theta}{2} \pm \sqrt{\frac{(1+\beta)^2\theta^2}{4} - \beta\theta}$$

2.4.1 Eigenvalues

Let

$$B := (1+\beta)I - sA, \quad C := I - tA$$

$$\begin{aligned} & Cy = \mu x, \quad -\beta x + By = \mu y \\ & \quad \Rightarrow \\ & -\beta Cy + \mu By = \mu^2 y \\ & \quad \Rightarrow \\ & \left(\mu - \frac{1}{2}B\right)^2 y = (\frac{1}{4}B^2 - \beta C) \end{aligned}$$

In the Nesterov-scheme we have $s = (1+\beta)t$ and $t = 1/\lambda_{\max}$, so $B = (1+\beta)C$ and for any $\lambda \in \sigma(A)$ with $\theta = \theta(\lambda) = 1 - \lambda/\lambda_{\max}$

$$\left(\mu - \frac{1+\beta}{2}\theta\right)^2 = \frac{(1+\beta)^2}{4}\theta^2 - \beta\theta$$

so

$$|\mu| = \begin{cases} \frac{1+\beta}{2}\theta + \sqrt{\frac{(1+\beta)^2}{4}\theta^2 - \beta\theta} & (1+\beta)^2\theta \geqslant 4\beta \\ \sqrt{\beta\theta} & \text{else} \end{cases}$$

With $\phi(x) = \alpha x + \sqrt{\alpha^2 x^2 - \beta x}$ we have $\phi'(x) = \alpha + \frac{2\alpha^2 x - \beta}{2\sqrt{\alpha^2 x^2 - \beta x}} \geqslant 0$ for $0 \leqslant x \leqslant 1$ if $\alpha^2 x^2 \geqslant \beta x$, so with $\rho = (1 - \lambda_{\min}/\lambda_{\max}) = \theta(\lambda_{\min})$

$$|\mu|(\lambda) = \begin{cases} \frac{1+\beta}{2}\rho + \frac{\sqrt{\rho}}{2}\sqrt{(1+\beta)^2\rho - 4\beta} & (1+\beta)^2\rho \geqslant 4\beta \\ \sqrt{\beta\rho} & \text{else} \end{cases}$$

Let $\rho = 1 - A^2 = (1-A)(1+A)$. Let $\beta := (1-A)/(1+A) = \rho/(1+A)^2$. Then

$$\frac{(1+\beta)^2\rho}{4\beta} = \frac{2(1+A)^2}{4(1+A)} = \frac{1+A}{2} \leqslant 1 \quad (A \leqslant 1),$$

so

$$|\mu|(\lambda) = \frac{1-A}{1+A}$$

Let now

$$\frac{(1+\beta)^2}{4\beta} = \rho^{-1}, \quad \mu = \sqrt{\beta\rho}$$

$$\begin{bmatrix} -\mu I & I - tA \\ -\beta & (1+\beta)(I-tA) - \mu I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

Then

$$\begin{aligned} (I - tA)y =& \mu x \\ ((1+\beta)(I-tA) - \mu I)y =& \beta x \end{aligned}$$

i.e.

$$\mu x = \rho y \quad \Leftrightarrow \quad x = \sqrt{\frac{\rho}{\beta}} y$$

2.4.2 Singular values

In order to bound the norm of the iteration matrix, we use the singular values, so

$$\begin{bmatrix} 0 & -\beta \\ I - tA & (1+\beta)I - sA \end{bmatrix} \begin{bmatrix} 0 & I - tA \\ -\beta & (1+\beta)I - sA \end{bmatrix} = \begin{bmatrix} \beta^2 & -\beta((1+\beta)I - sA) \\ -\beta((1+\beta)I - sA) & ((1+\beta)I - sA)^2 + (I - \end{bmatrix}$$

Let

$$\begin{bmatrix} \beta^2 I & -\beta((1+\beta)I - sA) \\ -\beta((1+\beta)I - sA) & ((1+\beta)I - sA)^2 + (I - tA)^2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mu^2 \begin{bmatrix} x \\ y \end{bmatrix}$$

or

$$\begin{bmatrix} \beta^2 I & -\beta B \\ -\beta B & B^2 + C^2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mu^2 \begin{bmatrix} x \\ y \end{bmatrix}, \quad B := (1+\beta)I - sA, \quad C := I - tA$$

so

$$(\mu^2-\beta^2)x = -\beta B y, \quad -\beta B x + (B^2+C^2)y = \mu^2 y$$
$$\Rightarrow$$
$$\beta^2 B^2 y + (\mu^2-\beta^2)(B^2+C^2)y = (\mu^2-\beta^2)\mu^2 y$$
$$\Rightarrow$$
$$(\mu^2-\beta^2)\mu^2 y = -\beta^2 C^2 y + \mu^2(B^2+C^2)y$$
$$\Rightarrow$$
$$\mu^4 - (\beta^2 I + B^2 + C^2)\mu^2 y = -\beta^2 C^2 y$$
$$\Rightarrow$$
$$\left(\mu^2 - \frac{\beta^2 I + B^2 + C^2}{2}\right)^2 y = \left(\frac{\beta^4 I + B^4 + C^4 + 2\beta^2 B^2 - 2\beta^2 C^2 + \beta^2 B^2 C^2}{4}\right) y$$

In the Nesterov-scheme we have $s = (1+\beta)t$ and $t = 1/\lambda_{\max}$, so $B = (1+\beta)C$ and

$$\left(\mu^2 - \frac{\beta^2 I + (2+\beta)C^2}{2}\right)^2 y = \left(\frac{\beta^4 I + (1+(1+\beta)^4+\beta^2(1+\beta)^2)C^4 + 2\beta^2((1+\beta)^2-1)C^2}{4}\right) y$$

with $\kappa = \lambda_{\max}/\lambda_{\min}$ and $\rho = 1 - 1/\kappa$

....

Lemma 4. *We have*

$$f(y_k) \leqslant f(x_k) \quad \Rightarrow \quad f(x_{k+1}) \leqslant f(x_k) \tag{19}$$

Proof. We have by hypothesis and convexity

$$f(x_{k+1}) \geqslant f(y_{k+1}) \geqslant f(x_{k+1}) + \beta(f(x_{k+1}) - f(x_k))$$

□

Lemma 5. *Suppose that*

$$\langle \nabla f(y_k), x_{k+1} - x_k \rangle \leqslant 0. \tag{20}$$

Proof. By the update rule (20) is equivalent to

$$\langle \nabla f(y_k), y_{k+1} - x_{k+1} \rangle \leqslant 0, \tag{21}$$

which gives with the update for x_{k+1}

$$\langle \nabla f(y_k), y_{k+1} - y_k \rangle \leqslant -t \left\|\nabla f(y_k)\right\|^2,$$

We have by convexity

$$
\begin{aligned}
f(y_k) \geqslant & f(y_{k+1}) + \langle \nabla f(y_{k+1}), y_k - y_{k+1} \rangle \\
= & f(y_{k+1}) + \langle \nabla f(y_k), y_k - y_{k+1} \rangle + \langle \nabla f(y_{k+1}) - \nabla f(y_k), y_k - y_{k+1} \rangle
\end{aligned}
$$

□

Proposition 2. *Suppose that* f *is* μ*-strongly convex and* ∇f *is* L*-Lipschitz. Then with* $\kappa_f := L/\mu$

$$??? \tag{22}$$

Proof. By the step-length rule we have

$$f(x_{k+1}) \leqslant f(y_k) - \frac{t_k}{2} \|\nabla f(y_k)\|^2 \tag{23}$$

By convexity we have

$$f(y_k) \leqslant f(x_k) + \langle \nabla f(y_k), y_k - x_k \rangle, \quad f(y_k) \leqslant f(x^*) + \langle \nabla f(y_k), y_k - x^* \rangle.$$

Convex combination with $0 \leqslant \alpha_k \leqslant 1$ gives

$$f(y_k) \leqslant \alpha_k f(x_k) + (1 - \alpha_k) f(x^*) + \langle \nabla f(y_k), y_k - \alpha_k x_k - (1 - \alpha_k) x^* \rangle$$

With (23), $\Delta f_k := f(x_k) - f(x^*)$ and the binomial identity $2ab - a^2 = b^2 - (b - a)^2$ we have

$$
\begin{aligned}
\Delta f_{k+1} - \alpha_k \Delta f_k = & f(x_{k+1}) - \alpha_k f(x_k) - (1 - \alpha_k) f(x^*) \\
\leqslant & \langle \nabla f(y_k), y_k - \alpha_k x_k - (1 - \alpha_k) x^* \rangle - \frac{t_k}{2} \|\nabla f(y_k)\|^2 \\
= & \frac{1}{2t_k} \left(\|y_k - \alpha_k x_k - (1 - \alpha_k) x^*\|^2 - \|y_k - \alpha_k x_k - (1 - \alpha_k) x^* - t_k \nabla f(y_k)\|^2 \right) \\
= & \frac{1}{2t_k} \left(\|y_k - \alpha_k x_k - (1 - \alpha_k) x^*\|^2 - \|x_{k+1} - \alpha_k x_k - (1 - \alpha_k) x^*\|^2 \right) \\
= & \frac{(1 - \alpha_k)^2}{2t_k} \left(\left\| \frac{y_k - \alpha_k x_k}{1 - \alpha_k} - x^* \right\|^2 - \left\| \frac{x_{k+1} - \alpha_k x_k}{1 - \alpha_k} - x^* \right\|^2 \right)
\end{aligned}
$$

Now we want

$$\frac{x_{k+1} - \alpha_k x_k}{1 - \alpha_k} = \frac{y_{k+1} - \alpha_{k+1} x_{k+1}}{1 - \alpha_{k+1}} \quad \Leftrightarrow \quad y_{k+1} = x_{k+1} + \beta_k (x_{k+1} - x_k), \quad \beta_k = \frac{\alpha_k (1 - \alpha_{k+1})}{1 - \alpha_k}$$

such that

$$\Delta f_{k+1} \leqslant \alpha_k \Delta f_k + \frac{(1-\alpha_k)^2}{2t_k}\left(\|x^* - \xi_k\|^2 - \|x^* + \xi_{k+1}\|^2\right), \quad \xi_k = y_k + \frac{\alpha_k}{1-\alpha_k}(y_k - x_k)$$

Or

$$\begin{aligned}(1+\alpha_k)\Delta f_{k+1} - \alpha_k \Delta f_k =& f(x_{k+1}) - f(x^*) + \alpha_k\left(f(x_{k+1}) - f(x_k)\right)\\ \leqslant& \langle \nabla f(y_k), (1+\alpha_k)y_k - \alpha_k x_k - x^* \rangle - \frac{t_k(1+\alpha_k)}{2}\|\nabla f(y_k)\|^2\\ =&\frac{(1+\alpha_k)}{2t_k}\left(2\langle t_k \nabla f(y_k), y_k - \frac{\alpha_k x_k + x^*}{(1+\alpha_k)}\rangle - \|t_k \nabla f(y_k)\|^2\right)\\ =&\frac{(1+\alpha_k)}{2t_k}\left(\left\|y_k - \frac{\alpha_k x_k + x^*}{(1+\alpha_k)}\right\|^2 - \left\|x_{k+1} - \frac{\alpha_k x_k + x^*}{(1+\alpha_k)}\right\|^2\right)\\ =&\frac{1}{2t_k(1+\alpha_k)}\left(\|(1+\alpha_k)y_k - \alpha_k x_k - x^*\|^2 - \|(1+\alpha_k)x_{k+1} - \alpha_k x_k - x\right.\end{aligned}$$

we want

$$(1+\alpha_k)y_k - \alpha_k x_k = (1+\alpha_{k-1})x_k - \alpha_{k-1}x_{k-1}$$

i.e.

$$y_k = x_k + \frac{\alpha_{k-1}}{1+\alpha_k}(x_k - x_{k-1})$$

Then

$$\Delta f_{k+1} \leqslant \alpha_k\left(\Delta f_k - \Delta f_{k+1}\right) + \frac{1}{2t_k(1+\alpha_k)}\left(\|x^* - \xi_k\|^2 - \|x^* - \xi_{k+1}\|^2\right)$$

with

$$\xi_k = x_k + \alpha_{k-1}(x_k - x_{k-1}) \tag{24}$$

$$\xi_k - y_k = \frac{\alpha_k \alpha_{k-1}}{1+\alpha_k}(x_k - x_{k-1}) = \alpha_k(y_k - x_k) = \frac{\alpha_k}{1+\alpha_k}(\xi_k - x_k)$$

Let

$$0 < \underline{\alpha} \leqslant \alpha_k \leqslant \overline{\alpha} < 1.$$

$$\|x^* - \xi_k\| \leqslant$$

Then we get

$$(1-\overline{\alpha})\sum_{k=n+1}^{\infty}\Delta f_k \leqslant \overline{\alpha}\Delta f_n + L(1-\underline{\alpha})^2\|x^* - \xi_n\|^2 \leqslant \overline{\alpha}\Delta f_n + 2\kappa_f(1-\underline{\alpha})^2\left(f(\xi_n) - f(x^*)\right)$$

.... □

If we impose

$$f(\xi_n) \leqslant f(x_n) \quad \Rightarrow \quad f(\xi_n) - f(x^*) \leqslant f(x_n) - f(x^*), \quad \underline{\beta} = \alpha\overline{\beta}$$

$$\sum_{k=n+1}^{\infty} \Delta f_k \leqslant \left(\frac{\overline{\beta} + 2\kappa_f (1 - \alpha\overline{\beta})^2}{1 - \overline{\beta}} \right) \Delta f_n$$

Let

$$\phi(s) := \frac{s + 2\kappa_f (1 - \alpha s)^2}{1 - s}, \quad \phi'(s) := \frac{(1 - 2\alpha s 2\kappa_f (1 - \alpha s))(1 - s) + \left(s + 2\kappa_f (1 - \alpha s)^2\right) 2s}{(1 - s)^2}$$

We have

$$Q(t_n, y_{n-1}, x_n) \leqslant Q(t_n, y_{n-1}, \xi_n) \quad \Rightarrow$$
$$\langle \nabla f(y_{n-1}), x_n - y_{n-1} \rangle + \frac{1}{2t_n} \|x_n - y_{n-1}\|^2 \leqslant \langle \nabla f(y_{n-1}), \xi_n - y_{n-1} \rangle + \frac{1}{2t_n} \|\xi_n - y_{n-1}\|^2 \quad \Rightarrow$$
$$0 \leqslant 2t_n \langle \nabla f(y_{n-1}), x_n - x_{n-1} \rangle + 2\langle x_n - y_{n-1}, x_n - x_{n-1} \rangle \; + \frac{\beta}{1 - \beta} \|x_n - x_{n-1}|$$