Nesterov's Accelerated Gradient Descent on α-strongly convex and β-smooth function

Proving NAGD converges at $\exp\left(-\dfrac{k-1}{\sqrt{Q}}\right)$

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft: August 2, 2017
Last update : July 29, 2019

Review : gradient descent

Given a α-strongly convex and β-smooth function $f : \mathbb{R}^n \to \mathbb{R}$, let $Q = \frac{\beta}{\alpha}$ be the conditional number.
On the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

Gradient Descent (GD) algorithm: starting with a inital point $x_0 \in \mathbb{R}^n$, iterates:

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

(if f is not differentiable, replace ∇f by sub-gradient)

GD takes $\mathcal{O}\left(Q \log \frac{1}{\epsilon}\right)$ iterations to reach ϵ-accuracy $\left(f(x_k) - f^* \leq \epsilon\right)$

Nesterov's Accelerated Gradient Descent (NAGD)

On the same problem, if f is β-smooth and convex, NAGD iterates

$$y_{k+1} = x_k - \frac{1}{\beta}\nabla f(x_k) \qquad x_{k+1} = (1-\gamma_k)y_{k+1} + \gamma_k y_k$$

$$\gamma_k = \frac{1-\lambda_k}{\lambda_{k+1}} \qquad \lambda_k = \frac{1+\sqrt{1+4\lambda_{k-1}^2}}{2}$$

If f is also α-strongly convex, NAGD iterates

$$y_{k+1} = x_k - \frac{1}{\beta}\nabla f(x_k) \qquad x_{k+1} = \left(1 - \frac{\sqrt{Q}-1}{\sqrt{Q}+1}\right) y_{k+1} + \frac{\sqrt{Q}-1}{\sqrt{Q}+1} y_k$$

Thus for strongly convex case $\gamma_k = \dfrac{\sqrt{Q}-1}{\sqrt{Q}+1}$

Convergence rate of NAGD on strongly convex and smooth function

Theorem The sequence $f(x_k)$ produced by NAGD on α-strongly convex and β-smooth function satisfies

$$f(y_k) - f^* \leq \frac{\alpha + \beta}{2} \|x_0 - x^*\|_2^2 \exp\left(-\frac{k-1}{\sqrt{Q}}\right)$$

The proof is long, it consists of proving a few lemmas.

Convergence rate of NAGD - the proof

Consider a functional $\Phi_k(x)$ that

- based on f
- is α-strongly convex
- has a general structure with "parameters" varies with iteration k.

In general $\Phi_k(x)$ can be defined as

$$\begin{aligned}
\Phi_0(x) &= f(x_0) + \frac{\alpha}{2}\|x - x_0\|_2^2 \\
\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x) + \frac{1}{\sqrt{Q}}\left(f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{\alpha}{2}\|x - x_k\|_2^2\right)
\end{aligned}$$

Φ_{k+1} is a convex combination of Φ_k and the 2nd order Taylor series approximation of f at x_k.
We have

$$\begin{aligned}
\nabla\Phi_0(x) &= \alpha(x - x_0) && (1) \\
\nabla^2\Phi_0(x) &= \alpha I_n && (2) \\
\nabla\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nabla\Phi_k(x) + \frac{1}{\sqrt{Q}}\nabla f(x_k) + \frac{\alpha}{\sqrt{Q}}(x - x_k) && (3) \\
\nabla^2\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nabla^2\Phi_k(x) + \frac{\alpha}{\sqrt{Q}}I_n && (4)
\end{aligned}$$

Lemma 1. $\Phi_{k+1}(x) \leq f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^k \Big(\Phi_0(x) - f(x)\Big)$, for all k.

Proof. By induction. For $k = 0$:

$$\Phi_{0+1}(x) = \Phi_1(x) = \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi(x) + \frac{1}{\sqrt{Q}}\Big(f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{\alpha}{2}\|x - x_0\|_2^2\Big)$$

f is α-strongly convex $f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{\alpha}{2}\|x - x_0\|_2^2 \leq f(x)$ so

$$\begin{aligned}
\Phi_1(x) &\leq \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi(x) + \frac{1}{\sqrt{Q}}f(x) \\
&= \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi(x) + \frac{1}{\sqrt{Q}}f(x) - f(x) + f(x) \\
&= f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^1 \Big(\Phi(x) - f(x)\Big)
\end{aligned}$$

so case $k = 0$ is proved.

Proving NAGD - proof lemma 1 ... 2/2

Induction Hypothesis $\Phi_k(x) \leq f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^{k-1} \left(\Phi(x_0) - f(x)\right)$

For case $k+1$

$$\Phi_{k+1}(x) = \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k(x) + \frac{1}{\sqrt{Q}} \left(f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{\alpha}{2} \|x - x_k\|_2^2\right)$$

f is α-strongly convex $f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{\alpha}{2} \|x - x_k\|_2^2 \leq f(x)$

$$\Phi_{k+1}(x) \leq \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k(x) + \frac{1}{\sqrt{Q}} f(x)$$

By induction hypothesis

$$\begin{aligned} \Phi_{k+1}(x) &\leq \left(1 - \frac{1}{\sqrt{Q}}\right) \left(f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^{k-1} \left(\Phi(x_0) - f(x)\right)\right) + \frac{1}{\sqrt{Q}} f(x) \\ &= f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^{k} \left(\Phi_0(x) - f(x)\right) \quad \square \end{aligned}$$

Lemma 2. $\nabla^2\Phi_k(x) = \alpha I_n$ where I_n is identity matrix.
Proof. By induction. For $k = 0$, by (2) $\nabla^2\Phi_0(x) = \alpha I_n$.

Induction Hypothesis $\nabla^2\Phi_k(x) = \alpha I_n$

For case $k+1$

$$\Phi_{k+1}(x) = \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x) + \frac{1}{\sqrt{Q}}\left(f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{\alpha}{2}\|x - x_k\|_2^2\right)$$

Take the Laplacian,

$$\begin{aligned}
\nabla^2\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nabla^2\Phi_k(x) + \frac{\alpha}{\sqrt{Q}}I_n \\
\text{[induction hypothesis]} &= \left(1 - \frac{1}{\sqrt{Q}}\right)\alpha I_n + \frac{\alpha}{\sqrt{Q}}I_n \\
&= \alpha I_n \quad \square
\end{aligned}$$

Proving NAGD - proof lemma 3 ... 1/10

Lemma 3. Let $\Phi_k^* = \min_{x \in \mathbb{R}^n} \Phi_k(x)$. Then $f(y_k) \leq \Phi_k^*$.

Proof. By induction. For $k = 0$, $\Phi_0^* = \Phi_0(x_0) = f(x_0) = f(y_0)$

Induction Hypothesis $f(y_k) \leq \Phi_k^*$

For case $k+1$: consider $f(y_{k+1})$ and β-smoothness of f

$$\begin{aligned}
f(y_{k+1}) &\leq f(x_k) + \nabla f(x_k)^T (y_{k+1} - x_k) + \frac{\beta}{2}\|y_{k+1} - x_k\|_2^2 \\
\text{[NAGD update]} &\leq f(x_k) + \nabla f(x_k)^T \left(\frac{-1}{\beta}\nabla f(x_k)\right) + \frac{\beta}{2}\|\frac{-1}{\beta}\nabla f(x_k)\|_2^2 \\
&= f(x_k) - \frac{1}{\beta}\|\nabla f(x_k)\|_2^2 + \frac{1}{2\beta}\|\nabla f(x_k)\|_2^2 \\
&= f(x_k) - \frac{1}{2\beta}\|\nabla f(x_k)\|_2^2
\end{aligned}$$

Let $g = \frac{1}{2\beta}\|\nabla f(x_k)\|_2^2$, we have $f(y_{k+1}) \leq f(x_k) - g$

Proving NAGD - proof lemma 3 ... 2/10

From $f(y_{k+1}) \leq f(x_k) - g$, two tricky steps to "create" $\left(1 - \frac{1}{\sqrt{Q}}\right)$

$$
\begin{array}{rcl}
f(y_{k+1}) & \leq & f(x_k) - \frac{f(x_k)}{\sqrt{Q}} + \frac{f(x_k)}{\sqrt{Q}} - g \\
& = & \left(1 - \frac{1}{\sqrt{Q}}\right) f(x_k) + \frac{f(x_k)}{\sqrt{Q}} - g \\
& = & \left(1 - \frac{1}{\sqrt{Q}}\right) f(x_k) - \left(1 - \frac{1}{\sqrt{Q}}\right) f(y_k) + \left(1 - \frac{1}{\sqrt{Q}}\right) f(y_k) + \frac{f(x_k)}{\sqrt{Q}} - g \\
& = & \left(1 - \frac{1}{\sqrt{Q}}\right) \Big(f(x_k) - f(y_k)\Big) + \left(1 - \frac{1}{\sqrt{Q}}\right) f(y_k) + \frac{f(x_k)}{\sqrt{Q}} - g \\
\begin{bmatrix}\text{induction} \\ \text{hypothesis}\end{bmatrix} & \leq & \left(1 - \frac{1}{\sqrt{Q}}\right) \Big(f(x_k) - f(y_k)\Big) + \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k^* + \frac{1}{\sqrt{Q}} f(x_k) - g \\
\left[f \text{ is cvx}^\dagger\right] & \leq & \left(1 - \frac{1}{\sqrt{Q}}\right) \nabla f(x_k)^T (x_k - y_k) + \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k^* + \frac{1}{\sqrt{Q}} f(x_k) - g
\end{array}
$$

We now have

$$
f(y_{k+1}) \quad \leq \quad \left(1 - \frac{1}{\sqrt{Q}}\right) \nabla f(x_k)^T (x_k - y_k) + \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k^* + \frac{1}{\sqrt{Q}} f(x_k) - g
$$

† f is convex so $f(x_y) - f(y_k) \leq \nabla f(x_k)^T (x_k - y_k)$.

We have

$$f(y_{k+1}) \quad \leq \quad \left(1 - \frac{1}{\sqrt{Q}}\right) \nabla f(x_k)^T (x_k - y_k) + \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k^* + \frac{1}{\sqrt{Q}} f(x_k) - g$$

Goal : to show

$$f(y_{k+1}) \leq \Phi_k^*$$

Hence we show

$$\left(1 - \frac{1}{\sqrt{Q}}\right) \nabla f(x_k)^T (x_k - y_k) + \left(1 - \frac{1}{\sqrt{Q}}\right) \Phi_k^* + \frac{1}{\sqrt{Q}} f(x_k) - g \leq \Phi_{k+1}^*$$

Now consider $\Phi_k(x)$. Lemma 2 $\nabla^2\Phi_k(x) = \alpha I_n$ implies

$$\Phi_k(x) = \Phi_k^* + \frac{\alpha}{2}\|x - \nu_k\|_2^2, \text{ for some } \nu_k \in \mathbb{R}^n$$

Note :

1. $\nabla\Phi_k(x) = \alpha(x - \nu_k)$
2. Φ_k is minimized at ν_k, which implies $\nabla\Phi_k(\nu_k) = 0$
3. Points (1),(2) work for all k: so also works for $k+1$
4. From $\Phi_0(x) = f(x_0) + \frac{\alpha}{2}\|x - x_0\|$, $\nu_0 = x_0$

By definition of $\Phi_{k+1}(x)$,

$$\begin{aligned}
\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x) + \frac{1}{\sqrt{Q}}\left(f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{\alpha}{2}\|x - x_k\|_2^2\right) \\
\nabla\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nabla\Phi_k(x) + \frac{\nabla f(x_k)}{\sqrt{Q}} + \frac{\alpha}{\sqrt{Q}}(x - x_k)
\end{aligned}$$

1. $\nabla\Phi_k(x) = \alpha(x - \nu_k)$
2. Φ_k is minimized at $\nu_k \iff \nabla\Phi_k(\nu_k) = 0$
3. Points (1),(2) work for all k
4. From $\Phi_0(x) = f(x_0) + \frac{\alpha}{2}\|x - x_0\|$, $\nu_0 = x_0$

What we have now

$$\begin{aligned}\nabla\Phi_{k+1}(x) &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nabla\Phi_k(x) + \frac{\nabla f(x_k)}{\sqrt{Q}} + \frac{\alpha}{\sqrt{Q}}(x - x_k)\\ \text{by (1)} &= \alpha\left(1 - \frac{1}{\sqrt{Q}}\right)(x - \nu_k) + \frac{\nabla f(x_k)}{\sqrt{Q}} + \frac{\alpha}{\sqrt{Q}}(x - x_k)\end{aligned}$$

Put $x = \nu_{k+1}$

$$\nabla\Phi_{k+1}(\nu_{k+1}) = \alpha\left(1 - \frac{1}{\sqrt{Q}}\right)(\nu_{k+1} - \nu_k) + \frac{\nabla f(x_k)}{\sqrt{Q}} + \frac{\alpha}{\sqrt{Q}}(\nu_{k+1} - x_k)$$

(2) + (3) : $\nabla\Phi_{k+1}(\nu_{k+1}) = 0$

$$\alpha\left(1 - \frac{1}{\sqrt{Q}}\right)(\nu_{k+1} - \nu_k) + \frac{\nabla f(x_k)}{\sqrt{Q}} + \frac{\alpha}{\sqrt{Q}}(\nu_{k+1} - x_k) = 0$$

From the last equality, re-arrange and express v_{k+1} as subject

$$\nu_{k+1} = \left(1 - \frac{1}{\sqrt{Q}}\right)\nu_k + \frac{1}{\sqrt{Q}}x_k - \frac{\nabla f(x_k)}{\alpha\sqrt{Q}} \tag{5}$$

From (5) we have

$$x_k - \nu_{k+1} = \left(1 - \frac{1}{\sqrt{Q}}\right)(x_k - \nu_k) + \frac{\nabla f(x_k)}{\alpha\sqrt{Q}} \tag{6}$$

Take norm-squared on (6)

$$\begin{aligned} \|x_k - \nu_{k+1}\|_2^2 = & \left(1 - \frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 \\ & + 2\left(1 - \frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k - \nu_k)}{\alpha\sqrt{Q}} \\ & + \frac{\|\nabla f(x_k)\|_2^2}{\alpha^2 Q} \end{aligned}$$

We have

$$\|x_k - \nu_{k+1}\|_2^2 = \left(1 - \frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 + 2\left(1 - \frac{1}{\sqrt{Q}}\right) \frac{\nabla f(x_k)^T(x_k - \nu_k)}{\alpha\sqrt{Q}} + \frac{\|\nabla f(x_k)\|_2^2}{\alpha^2 Q}$$

Now consider $\Phi_{k+1}(x)$ evaluate at x_k, from page 12 we have

$$\Phi_{k+1}(x_k) = \Phi_{k+1}^* + \frac{\alpha}{2}\|x_k - \nu_{k+1}\|_2^2$$

Which is

$$\Phi_{k+1}(x_k) = \Phi_{k+1}^* + \frac{\alpha}{2}\left(1 - \frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 + \left(1 - \frac{1}{\sqrt{Q}}\right) \frac{\nabla f(x_k)^T(x_k - \nu_k)}{\sqrt{Q}} + g$$

by using the fact $\alpha Q = \beta$ and $g = \frac{1}{2\beta}\|\nabla f(x_k)\|_2^2$.

Now we have an expression of $\Phi_{k+1}(x_k)$

$$\Phi_{k+1}(x_k) = \Phi^*_{k+1} + \frac{\alpha}{2}\left(1 - \frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 + \left(1 - \frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k - \nu_k)}{\sqrt{Q}} + g$$

By definition of $\Phi_{k+1}(x)$ from page 5, $\Phi_{k+1}(x_k)$ is

$$\Phi_{k+1}(x_k) = \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x_k) + \frac{1}{\sqrt{Q}}\left(f(x_k) + \nabla f(x_k)^T(x_k - x_k) + \frac{\alpha}{2}\|x_k - x_k\|_2^2\right)$$

Remove those $x_k - x_k$ terms

$$\Phi_{k+1}(x_k) = \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x_k) + \frac{f(x_k)}{\sqrt{Q}}$$

Equate the two expressions

$$\begin{aligned}\Phi^*_{k+1} = \quad & -\frac{\alpha}{2}\left(1 - \frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 - \left(1 - \frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k - \nu_k)}{\sqrt{Q}} - g \\ & + \left(1 - \frac{1}{\sqrt{Q}}\right)\Phi_k(x_k) + \frac{f(x_k)}{\sqrt{Q}}\end{aligned}$$

continue from next page

Proving NAGD - proof lemma 3 ... 9/10

$$\Phi_{k+1}^* = -\frac{\alpha}{2}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k-\nu_k\|_2^2 - \left(1-\frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k-\nu_k)}{\sqrt{Q}} - g + \left(1-\frac{1}{\sqrt{Q}}\right)\Phi_k(x_k) + \frac{f(x_k)}{\sqrt{Q}}$$

By $\Phi_k(x) = \Phi_k^* + \frac{\alpha}{2}\|x-\nu_k\|_2^2$ (page 12)

$$\left(1-\frac{1}{\sqrt{Q}}\right)\Phi_k(x_k) = \left(1-\frac{1}{\sqrt{Q}}\right)\Phi_k^* + \left(1-\frac{1}{\sqrt{Q}}\right)\frac{\alpha}{2}\|x_k-\nu_k\|_2^2$$

Hence

$$\Phi_{k+1}^* = \underbrace{-\frac{\alpha}{2}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k-\nu_k\|_2^2}_{*} - \left(1-\frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k-\nu_k)}{\sqrt{Q}} - g + \left(1-\frac{1}{\sqrt{Q}}\right)\Phi_k^* + \underbrace{\left(1-\frac{1}{\sqrt{Q}}\right)\frac{\alpha}{2}\|x_k-\nu_k\|_2^2}_{*} + \frac{f(x_k)}{\sqrt{Q}}$$

Simplify the *

$$\Phi_{k+1}^* = \left(1-\frac{1}{\sqrt{Q}}\right)\Phi_k^* + \frac{\alpha}{2\sqrt{Q}}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k-\nu_k\|_2^2 - \left(1-\frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k-\nu_k)}{\sqrt{Q}} - g + \frac{f(x_k)}{\sqrt{Q}}$$

To proceed, we need lemma 4.

Lemma 4. $\nu_k - x_k = \sqrt{Q}(x_k - y_k)$.

Proof by induction. For $k = 0$, as $\nu_0 = x_0$ and $x_0 = y_0$,

$$\nu_0 - x_0 = 0 = \sqrt{Q}(x_0 - y_0)$$

Induction hypothesis $\nu_k - x_k = \sqrt{Q}(x_k - y_k)$

For case $k + 1$

$$\begin{aligned}
\nu_{k+1} &\overset{(5)}{=} \left(1 - \frac{1}{\sqrt{Q}}\right)\nu_k + \frac{x_k}{\sqrt{Q}} - \frac{\nabla f(x_k)}{\alpha\sqrt{Q}} \\
\left[\frac{1}{\alpha\sqrt{Q}} = \frac{\sqrt{Q}}{\beta}\right] &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nu_k + \frac{x_k}{\sqrt{Q}} - \frac{\sqrt{Q}\nabla f(x_k)}{\beta} \\
\nu_{k+1} - x_{k+1} &= \left(1 - \frac{1}{\sqrt{Q}}\right)\nu_k + \frac{x_k}{\sqrt{Q}} - \frac{\sqrt{Q}\nabla f(x_k)}{\beta} - x_{k+1} \\
\text{[induction hypothesis]} &= \sqrt{Q}x_k - \left(\sqrt{Q} - 1\right)y_k - \frac{\sqrt{Q}\nabla f(x_k)}{\beta} - x_{k+1} \\
&= \sqrt{Q}\left(x_k - \frac{\nabla f(x_k)}{\beta}\right) - \left(\sqrt{Q} - 1\right)y_k - x_{k+1} \\
\text{by (1)} &= \sqrt{Q}y_{k+1} - \left(\sqrt{Q} - 1\right)y_k - x_{k+1}
\end{aligned}$$

Proving NAGD - proof lemma 4 ... 2/2

Now we have $\nu_{k+1} - x_{k+1} = \sqrt{Q}y_{k+1} - (\sqrt{Q}-1)y_k - x_{k+1}$.

By (2), $x_{k+1} = \left(1 + \frac{\sqrt{Q}-1}{\sqrt{Q}+1}\right) y_{k+1} - \frac{\sqrt{Q}-1}{\sqrt{Q}+1} y_k$, re-arrange gives

$$-(\sqrt{Q}-1)y_k \quad = \quad (\sqrt{Q}+1)x_{k+1} - 2\sqrt{Q}y_{k+1}$$

Therefore

$$\begin{aligned} \nu_{k+1} - x_{k+1} &= \sqrt{Q}y_{k+1} + (\sqrt{Q}+1)x_{k+1} - 2\sqrt{Q}y_{k+1} - x_{k+1} \\ &= \sqrt{Q}(x_{k+1} - y_{k+1}) \quad \square \end{aligned}$$

We can now back to lemma 3

The proof of Lemma 3 stops at

$$\begin{aligned}\Phi^*_{k+1} = \;& \left(1-\frac{1}{\sqrt{Q}}\right)\Phi^*_k + \frac{\alpha}{2\sqrt{Q}}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k - \nu_k\|_2^2 \\ & - \left(1-\frac{1}{\sqrt{Q}}\right)\frac{\nabla f(x_k)^T(x_k-\nu_k)}{\sqrt{Q}} - g + \frac{f(x_k)}{\sqrt{Q}}\end{aligned}$$

Apply lemma 4 $\nu_k - x_k = \sqrt{Q}(x_k - y_k)$, we have

$$\begin{aligned}\Phi^*_{k+1} = \;& \left(1-\frac{1}{\sqrt{Q}}\right)\Phi^*_k + \frac{\alpha\sqrt{Q}}{2}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k - y_k\|_2^2 \\ & + \left(1-\frac{1}{\sqrt{Q}}\right)\nabla f(x_k)^T(x_k-y_k) - g + \frac{f(x_k)}{\sqrt{Q}}\end{aligned}$$

Recall (page 11)

$$f(y_{k+1}) \quad\leq\quad \underbrace{\left(1-\frac{1}{\sqrt{Q}}\right)\nabla f(x_k)^T(x_k-y_k) + \left(1-\frac{1}{\sqrt{Q}}\right)\Phi^*_k + \frac{f(x_k)}{\sqrt{Q}} - g}_{F}$$

Thus

$$\Phi_{k+1} = \frac{\alpha\sqrt{Q}}{2}\left(1-\frac{1}{\sqrt{Q}}\right)^2 \|x_k - y_k\|_2^2 + F \geq F \geq f(y_{k+1}). \quad \square$$

Proving NAGD - prove the whole theorem

Combine lemma 1 and lemma 3 proves the theorem.

Want to prove : $f(y_k) - f^* \leq \frac{\alpha+\beta}{2}\|x_0 - x^*\|_2^2 \exp\left(-\frac{k-1}{\sqrt{Q}}\right)$

Lemma 1. $\Phi_{k+1}(x) \leq f(x) + \left(1 - \frac{1}{\sqrt{Q}}\right)^k \left(\Phi_0(x) - f(x)\right)$, for all k.

Lemma 3. $f(y_k) \leq \Phi_k^*$.

Also f is β-smooth and $\nabla f(x)$ vanish at x^* so

$f(x_0) - f(x^*) \leq \nabla f(x^*)^T(x_0 - x^*) + \frac{\beta}{2}\|x_0 - x^*\|_2^2 = \frac{\beta}{2}\|x_0 - x^*\|_2^2$

The proof.

$$
\begin{array}{rcll}
f(y_k) - f^* & \leq & \Phi_k(x^*) - f^* & \text{lemma 3} \\
& \leq & f(x^*) + \left(1 - \frac{1}{\sqrt{Q}}\right)^{k-1} \left(\Phi_0(x^*) - f(x^*)\right) - f^* & \text{lemma 1} \\
& = & \left(\Phi_0(x^*) - f^*\right)\left(1 - \frac{1}{\sqrt{Q}}\right)^{k-1} & f(x^*) = f^* \\
& \leq & \left(\Phi_0(x^*) - f^*\right)\exp\left(-\frac{k-1}{\sqrt{Q}}\right) & 1 + x \leq e^x \\
& = & \left(f(x_0) - f^* + \frac{\alpha}{2}\|x_0 - x^*\|_2^2\right)\exp\left(-\frac{k-1}{\sqrt{Q}}\right) & \text{Def. of } \Phi_0(x) \\
& \leq & \left(\frac{\beta}{2}\|x_0 - x^*\|_2^2 + \frac{\alpha}{2}\|x_0 - x^*\|_2^2\right)\exp\left(-\frac{k-1}{\sqrt{Q}}\right) & f \text{ is } \beta\text{-smooth} \\
& \leq & \frac{\alpha+\beta}{2}\|x_0 - x^*\|_2^2 \exp\left(-\frac{k-1}{\sqrt{Q}}\right) \quad \square &
\end{array}
$$

Discussion

If we stop the algorithm when ϵ-accuracy is achieved

$$\frac{\alpha+\beta}{2}\|x_0-x^*\|_2^2 \exp\left(-\frac{k-1}{\sqrt{Q}}\right) \leq \epsilon$$

Re-arrange

$$k \geq \sqrt{Q}\ln\frac{1}{\epsilon} + \text{constant}$$

i.e. it takes $\mathcal{O}\left(\sqrt{Q}\ln\frac{1}{\epsilon}\right)$ steps for NAGD to converges. Compared to GD with rate $\mathcal{O}\left(Q\ln\frac{1}{\epsilon}\right)$, the improvement $Q \to \sqrt{Q}$ is significant as α can be viewed as regularization parameter in various machine learning model (norm regularized) and $\frac{1}{\alpha}$ can be as large as sample size. Here the number of step reduced from sample size to $\sqrt{\text{sample size}}$.

Last page - summary

For unconstrained problem $\min_{x \in \mathbb{R}^n} f(x)$, with $f : \mathbb{R}^n \to \mathbb{R}$ being β-smooth and α-strongly convex, the NAGD algorithm iterates the following :

$$y_{k+1} = x_k - \frac{1}{\beta} \nabla f(x_k) \qquad x_{k+1} = \left(1 - \frac{\sqrt{Q}-1}{\sqrt{Q}+1}\right) y_{k+1} + \frac{\sqrt{Q}-1}{\sqrt{Q}+1} y_k$$

with initial point $x_0 = y_0 \in \mathbb{R}^n$, will produce a sequences $f(y_k)$ that

$$f(y_k) - f^* \leq \frac{\alpha+\beta}{2} \|x_0 - x^*\|_2^2 \exp\left(-\frac{k-1}{\sqrt{Q}}\right)$$

End of document