Gradient methods

Roland Becker

January 31, 2023

Contents

1 Class of methods

Lemma 1. *Let* $q_i(x) = q_i^* + \frac{\mu}{2}\|x - v_i\|^2$. *Then for* $\lambda \in [0,1]$ *we have for* $q_\lambda := (1-\lambda)q_0 + \lambda q_1$ *the expression*

$$(1-\lambda)q_0(x) + \lambda q_1(x) = q_\lambda^* + \frac{\mu}{2}\|x - v_\lambda\|^2 \tag{1}$$

with

$$v_\lambda = (1-\lambda)v_0 + \lambda v_1, \quad q_\lambda^* = (1-\lambda)q_0^* + \lambda q_1^* + \frac{\lambda(1-\lambda)\mu}{2}\|v_1 - v_0\|^2. \tag{2}$$

Furthermore

$$\operatorname{argmin}\{q_\lambda^* \mid 0 \leqslant \lambda \leqslant 1\} = P_{[0,1]}\left(\frac{1}{2} + \frac{q_1^* - q_0^*}{\mu\|v_1 - v_0\|^2}\right). \tag{3}$$

Proof. The two functions in (1) are quadratic un x and λ. Since they coincide for $\lambda = 0$ and $\lambda = 1$, it is sufficient to check that the second derivative with respect to λ of the right-hand side vanishes:

$$\frac{\partial^2}{\partial\lambda^2}\|x - v_\lambda\|^2 = 2\|v_1 - v_0\|^2 = -\frac{\partial^2\lambda(1-\lambda)}{\partial\lambda^2}\|v_1 - v_0\|^2$$

□

For $f \in \mathcal{F}_{\mu,L}(X)$ we have

$$f(x) \geqslant f(y) + \langle\nabla f(y), x - y\rangle + \frac{\mu}{2}\|x - y\|^2 = f(y) - \frac{1}{2\mu}\|\nabla f(y)\|^2 + \frac{\mu}{2}\left\|x - y + \frac{1}{\mu}\nabla f(y)\right\|^2$$

Algorithm 1: GM fixed step size

Input: $x_0 \in X$, $\varepsilon > 0$.

Set $v_0 = x_0$, $q_0 = -\infty$ and $k = 0$.

(1) If $f(x_k) - q_k \leqslant \varepsilon$: STOP

(5) $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$

(3) $w_k = x_k - \frac{1}{\mu}\nabla f(x_k)$, $p_k = f(x_k) - \frac{\|\nabla f(x_k)\|^2}{2\mu}$

$Q_k = \frac{\mu}{2}\|v_k - w_k\|^2$, $\lambda_k = P_{[0,1]}\left(\frac{1}{2} + \frac{p_k - q_k}{2Q_k}\right)$

$q_{k+1} = (1-\lambda_k)q_k + \lambda_k p_k + \lambda_k(1-\lambda_k)Q_k$

(4) $v_{k+1} = (1-\lambda_k)v_k + \lambda_k w_k$

(6) Increment k and go to (1)

Algorithm 2: AGM fixed step size

Input: $x_0 \in X$, $\varepsilon > 0$.

Set $v_0 = x_0 - \frac{1}{\mu}\nabla f(x_0)$, $q_0 = f(x_0) - \frac{\|\nabla f(x_0)\|^2}{2\mu}$ and $k = 0$.

(1) If $f(x_k) - q_k \leqslant \varepsilon$: STOP

(2) $y_k = (1-\theta_k)v_k + \theta_k x_k$

(3) $Q_k = \frac{\mu}{2}\left\|v_k - y_k + \frac{1}{\mu}\nabla f(y_k)\right\|^2$

$q_{k+1} = (1-\lambda_k)q_k + \lambda_k\left(f(y_k) - \frac{\|\nabla f(y_k)\|^2}{2\mu}\right) + \lambda_k(1-\lambda_k)Q_k$

(4) $v_{k+1} = (1-\lambda_k)v_k + \lambda_k\left(y_k - \frac{1}{\mu}\nabla f(y_k)\right)$

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

(6) Increment k and go to (1)

For $\theta_k < 1$ we have

$$v_k - y_k = \frac{\theta_k}{1-\theta_k}(y_k - x_k)$$

such that

$$Q_k = \frac{\theta_k^2\mu}{2(1-\theta_k)^2}\left\|x_k - y_k\right\|^2 + \frac{1}{2\mu}\left\|\nabla f(y_k)\right\|^2 + \frac{\theta_k}{1-\theta_k}\langle \nabla f(y_k), y_k - x_k\rangle.$$

Let $e_k := f(x_k) - q_k$ Then we have

$$\begin{aligned} e_{k+1} \leqslant & f(y_k) - \frac{1}{2L}\left\|\nabla f(y_k)\right\|^2 - q_{k+1} \qquad (5)\\ =&(1-\lambda_k)(f(y_k)-q_k) - \frac{1}{2L}\left\|\nabla f(y_k)\right\|^2 + \lambda_k\frac{\left\|\nabla f(y_k)\right\|^2}{2\mu} - \lambda_k(1-\lambda_k)Q_k\\ =&(1-\lambda_k)e_k + \underbrace{(1-\lambda_k)\langle \nabla f(y_k), y_k - x_k\rangle + \left(\frac{\lambda_k}{2\mu} - \frac{1}{2L}\right)\left\|\nabla f(y_k)\right\|^2 - \lambda_k(1-\lambda_k)Q_k}_{=:X} \end{aligned}$$

$$\begin{aligned} X =&(1-\lambda_k)\left(1 - \frac{\lambda_k\theta_k}{1-\theta_k}\right)\langle \nabla f(y_k), y_k - x_k\rangle + \left(\frac{\lambda_k}{2\mu} - \frac{1}{2L} - \frac{\lambda_k(1-\lambda_k)}{2\mu}\right)\left\|\nabla f(y_k)\right\|^2\\ &- \frac{\lambda_k(1-\lambda_k)\theta_k^2\mu}{2(1-\theta_k)^2}\left\|x_k - y_k\right\|^2\\ =&\frac{(1-\lambda_k)(1-(1+\lambda_k)\theta_k)}{1-\theta_k}\langle \nabla f(y_k), y_k - x_k\rangle + \left(\frac{\lambda_k^2}{2\mu} - \frac{1}{2L}\right)\left\|\nabla f(y_k)\right\|^2\\ &- \frac{\lambda_k(1-\lambda_k)\theta_k^2\mu}{2(1-\theta_k)^2}\left\|x_k - y_k\right\|^2 \end{aligned}$$

Under the condition $(1+\lambda_k)\theta_k < 1$ we get

$$X \leqslant \left(\frac{2(1-\theta_k)^2}{4\lambda_k(1-\lambda_k)\theta_k^2\mu}\frac{(1-\theta_k)^2}{(1-\lambda_k)^2(1-(1+\lambda_k)\theta_k)^2} + \frac{\lambda_k^2}{2\mu} - \frac{1}{2L}\right)\left\|\nabla f(y_k)\right\|^2$$