Report for 2024/07/30

Hiroki Hamaguchi

I. Introduction

problem:

$$\min f(x) \tag{I.1}$$

descent direction d_k:

$$d_k = \begin{cases} -g_k & \text{for } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{for } k > 1 \end{cases} \tag{I.2}$$

update rule:

$$x_{k+1} = x_k + \alpha_k d_k \tag{I.3}$$

FR method (erratic):

$$\beta_k^{\mathrm{FR}} = \|g_k\|^2 / \|g_{k-1}\|^2 \tag{I.4}$$

PR method (preferred) :

$$\beta_k^{\mathrm{PR}} = \langle g_k, g_k - g_{k-1} \rangle / \|g_{k-1}\|^2 \tag{I.5}$$

HS method (similar to PR) :

$$\beta_k^{\mathrm{HS}} = \langle g_k, g_k - g_{k-1} \rangle / \langle d_{k-1}, g_k - g_{k-1} \rangle \tag{I.6}$$

II. Preliminaries

$s_k := x_{k+1} - x_k$
$y_k := g_{k+1} - g_k$
d_k is a descent direction if $\langle g_k, d_k \rangle < 0$.
angle between $-g_k$ and d_k:

$$\cos\theta_k := -\langle g_k, d_k \rangle / \|g_k\| \|d_k\| \tag{II.1}$$

Assumption 1. *1) the level set $\mathcal{L} := \{x \mid f(x) \leq f(x_1)\}$ is bounded.*
2) there exists some neighborhood $\mathcal{N}$ of $\mathcal{L}$ such that, for all $x, \overline{x} \in \mathcal{N}$ there exists $L > 0$ such that

$$\|g(x) - g(\overline{x})\| \leq L \|x - \overline{x}\|. \tag{II.2}$$

This implies there exists a constant $\overline{\gamma}$ such that

$$\|g_k\| \leq \overline{\gamma}, \quad \text{for all } x \in \mathcal{L}. \tag{II.3}$$

Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k \langle g_k, d_k \rangle \tag{II.4}$$

$$\langle g(x_k + \alpha_k d_k), d_k \rangle \geq \sigma_2 \langle g_k, d_k \rangle \tag{II.5}$$

where $0 < \sigma_1 < \sigma_2 < 1$.
More ideal condition:

$$f(x_k + \alpha_k d_k) \leq f(x_k + \hat{\alpha}_k d_k) \tag{II.6}$$

where $\hat{\alpha}_k$ is the smallest positive stationary point of the function $\xi_k(\alpha) := f(x_k + \alpha d_k)$.

Theorem 1. *Suppose that Assumption 1 holds, and α_k satisfies the Wolfe conditions or the more ideal condition (II.6). Then, Zoutendijk condition holds:*

$$\sum_{k \leq 1} \cos^2\theta_k \|g_k\|^2 < \infty. \tag{II.7}$$

exact line search (orthogonality condition):

$$\langle g_k, d_{k-1} \rangle = 0 \tag{II.8}$$

By (I.2), we have

$$\cos\theta_k = \frac{\|g_k\|}{\|d_k\|} \tag{II.9}$$

Substituting this in Zoutendijk condition, we have

$$\sum_{k \leq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \tag{II.10}$$

If $\{\|d_k\| / \|g_k\|\}$ is bounded, then

$$\lim_{k \to \infty} \|g_k\| = 0 \tag{II.11}$$

However, this is not always the case. We can only obtain a weaker result:

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{II.12}$$

If (II.12) does not hold, then there exists $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \quad \text{for all } k \geq 1. \tag{II.13}$$

and this implies

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty. \tag{II.14}$$

By Zoutendijk, if (II.13) holds, then

$$\|d_k\|^2 \leq ck,$$

for some constant c.
It contradicts (II.14), so (II.13) does not hold.
For inexact, we can proceed if

$$\cos\theta_k \geq c \|g_k\| / \|d_k\|, \tag{II.15}$$

for some constant c. The rest of the analysis is the same.
strong Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k \langle g_k, d_k \rangle \tag{II.16}$$

$$|\langle g(x_k + \alpha_k d_k), d_k\rangle| \geq -\sigma_2 \langle g_k, d_k\rangle \tag{II.17}$$

where $0 < \sigma_1 < \sigma_2 < 1$. If strong Wolfe conditions hold, then FR method gives (II.15).

(II.15) is equivalent to

$$\langle g_k, d_k\rangle \leq -c\|g_k\|^2. \tag{II.18}$$

$$\beta_k = \max \beta_k^{\mathrm{PR}}, 0 \tag{II.19}$$

III. FR METHOD

definition of β_k: any scalar such that

$$|\beta_k| \leq \beta_k^{\mathrm{FR}} \tag{III.1}$$

for all $k \geq 2$.

Lemma 1. *Suppose that Assumption 1 holds.* (III.1) *with strong Wolfe condition implies*

$$-\frac{1}{1-\sigma_2} \leq \frac{\langle g_k, d_k\rangle}{\|g_k\|^2} \leq \frac{2\sigma_2 - 1}{1-\sigma_2} \tag{III.2}$$

for all $k \geq 1$.

Proof. Since $0 < \sigma_2 < \frac{1}{2}$,

$$\frac{2\sigma_2 - 1}{1-\sigma_2} < 0. \tag{III.3}$$

$$\begin{aligned} \frac{\langle g_{k+1}, d_{k+1}\rangle}{\|g_{k+1}\|^2} &= -1 + \beta_{k+1}\frac{\langle g_{k+1}, d_k\rangle}{\|g_{k+1}\|^2} \\ &= -1 + \frac{\beta_{k+1}}{\beta_{k+1}^{\mathrm{FR}}}\frac{\langle g_{k+1}, d_k\rangle}{\|g_k\|^2} \end{aligned} \tag{III.4}$$

$$|\beta_{k+1}\langle g_{k+1}, d_k\rangle| \leq -\sigma_2|\beta_{k+1}|\langle g_k, d_k\rangle$$

TO DO ∎

IV. RELATED TO PR

We require the sufficient decrease condition:

$$\langle g_k, d_k\rangle \leq -\sigma_3\|g_k\|^2 \tag{IV.1}$$

for some constant $0 < \sigma_3 \leq 1$ and for all $k \geq 1$.

From (I.2),

$$\langle g_k, d_k\rangle = -\|g_k\|^2 + \beta_k\langle g_k, d_{k-1}\rangle \tag{IV.2}$$

If $\langle g_k, d_{k-1}\rangle \leq 0$, then the non-negativity of β_k implies (IV.1). If (IV.1) does not hold, then $\langle g_k, d_{k-1}\rangle > 0$.

Proof by contradiction, assuming:

$$\text{for some } \gamma > 0, \quad \|g_k\| \geq \gamma \text{ for all } k \geq 1. \tag{IV.3}$$