Exercice 4 - Affine-invariance of the BFGS algorithm

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Frechet differentiable objective function. Consider the BFGS algorithm defined as

initialize state $\theta_0 = (x_0, B_0) \in \mathbb{R}^n \times \mathcal{S}_{n,>}(\mathbb{R})$, $k = 0$
while stopping criterion not met **do**
  compute $d_k = -B_k^{-1} \nabla f(x_k)$
  compute step-size: $\alpha_k = \text{LineSearch}(x_k, d_k, f)$
  move in the direction of d_k: $x_{k+1} = x_k + \alpha_k d_k$
  compute $s_k = \alpha_k d_k$
  compute $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$
  update estimate of Hessian: $B_{k+1} = B_k + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k}$
  $k = k + 1$
end while

We will assume for the sake of simplicity that the step-size $\alpha_k = \alpha$ is constant (the proof works also with optimal step-size).

Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix and let $x_0 \in \mathbb{R}^n$ and $B_0 \in \mathbb{R}^{n\times n}$ with $B_0 \succ 0$. Consider the sequence $(x_k, B_k)_{k\geq 1}$ generated by the BFGS algorithm optimizing $x \mapsto f(x)$. Let $(x'_0, B'_0) = (A^{-1}x_0, A^T B_0 A)$ and consider $(x'_k, B'_k)_{k\geq 1}$ the sequence of states of the BFGS algorithm optimizing $g(x') = f(Ax')$ and initialized in (x'_0, B'_0).

Prove that for all $k \geq 1$, $(x'_k, B'_k) = (A^{-1}x_k, A^T B_k A)$, i.e. that the BFGS algorithm is affine-invariant.

Solution:

Let $\{(x_t, B_t) : t \geq 0\}$ be the sequence of state of the BFGS algorithm optimizing f and similarly $\{(x'_t, B'_t) : t \geq 0\}$ be the sequence of states optimizing $g(x) = f(Ax + b)$.

Assume that $(x'_t, B'_t) = (A^{-1}x_t, A^T B_t A)$, we need to show that that $(x'_{t+1}, B'_{t+1}) = (A^{-1}x_{t+1}, AB_{t+1}A)$.

By definition $x'_{t+1} = x'_t + \alpha(-[B'_t]^{-1})\nabla g(x'_t) = x'_t + \alpha(-[B'_t]^{-1})A^T\nabla f(Ax'_t + b) = A^{-1}(x_t - b) + \alpha(-A^{-1}[B_t]^{-1}A^{-T})A^T\nabla f(Ax'_t + b) = A^{-1}(x_t - \alpha[B_t]^{-1}\nabla f(x_t) - b)$. Since $x_{t+1} = x_t - \alpha[B_t]^{-1}\nabla f(x_t)$, we find that $x'_{t+1} = A^{-1}(x_{t+1} - b)$.
Similarly, we show that $B'_{t+1} = A^T B_{t+1} A$. Start from

$$B'_{t+1} = B'_t + \frac{(\nabla g(x'_{t+1}) - \nabla g(x'_t))(\nabla g(x'_{t+1}) - \nabla g(x'_t))^T}{(\nabla g(x'_{t+1}) - \nabla g(x'_t))^T \alpha(-B'_t)^{-1}\nabla g(x'_t)} - \frac{B'_t s s^T B'_t}{s^T B'_t s}$$

where $s = \alpha p$. We do in details the computation for the middle term. Since $\nabla g(x) = A^T\nabla f(Ax + b)$ we find that $\nabla g(x'_{t+1}) - \nabla g(x'_t) = A^T[\nabla f(Ax'_{t+1} + b) - \nabla f(Ax'_t + b)] = A^T[\nabla f(x_{t+1}) - \nabla f(x_t)]$ and thus

$$\frac{(\nabla g(x'_{t+1}) - \nabla g(x'_t))(\nabla g(x'_{t+1}) - \nabla g(x'_t))^T}{(\nabla g(x'_{t+1}) - \nabla g(x'_t))^T \alpha(-B'_t)^{-1}\nabla g(x'_t)} = A^T \frac{(\nabla f(x_{t+1}) - \nabla f(x_t))(\nabla f(x_{t+1}) - \nabla f(x_t))^T}{(\nabla f(x_{t+1}) - \nabla f(x_t))^T \alpha(-B_t)^{-1}\nabla f(x_t)} A \quad (8)$$

that is the middle term of the update for B'_{t+1} is A^T times the middle term of the update for B_{t+1} times A. The computation for the rightmost term works in the same way.