R. Herzog, G. Müller, E. Herberg, M. Hashemi
Heidelberg University

Nonlinear Optimization
Spring Semester 2023

Exercise 6 - Solution

Date issued: 22nd May 2023
Date due: 31st May 2023

Homework Problem 6.1 (Truncated Newton CG) 6 Points

Implement the truncated Newton-CG method (Algorithm 5.44 with Algorithm 5.41), apply it for Rosenbrock's and/or Himmelblau's functions and compare its performance with the exact globalized Newton method.

Solution.

For the implementation see the file driver_ex_021_compare_vanilla_inexact_newton.py. The modifications compared to the globalized Newton method implemented in homework problem 5.4 are minor. Simply replace whatever linear system solver was in place until that point by the CG method, start at the zero initial guess and only use relative tolerances generated by the driving sequence.

The results in Figure 0.1 show that in early iterations, both methods obviously differ, but the inexact solution does not necessarily take worse steps in the beginning (though it sometimes does). The overall convergence behavior is similar though, with superlinear/quadratic convergence being observable. The initial step length suggestion of 1 is typically accepted after a few iterations. Both methods neither detect nonpositive curvature nor does the computed direction violate the generalized angle condition in this example.

The results in Figure 0.2 show that in early iterations, both methods obviously differ, which can lead to convergence to different local minima. The overall convergence speed behaves similarly though, with superlinear/quadratic convergence being observable. The initial step length suggestion of 1 is also typically accepted after a few iterations. In this example, both the globalized Newton as well as

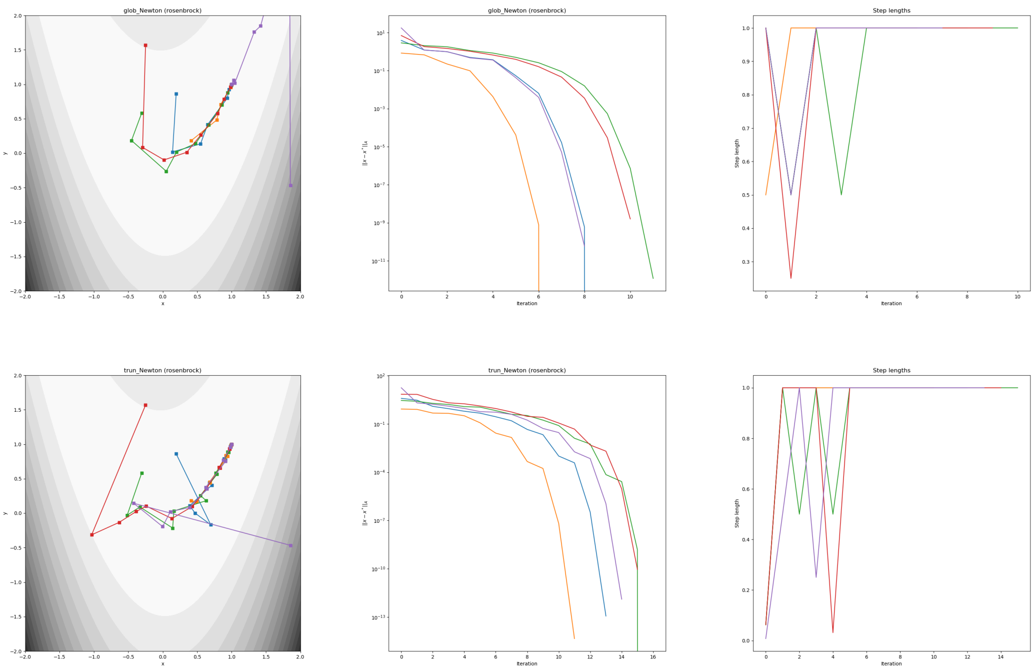

Figure 0.1: Exact Newton (top) vs truncated Newton CG (bottom) for Rosenbrock function.

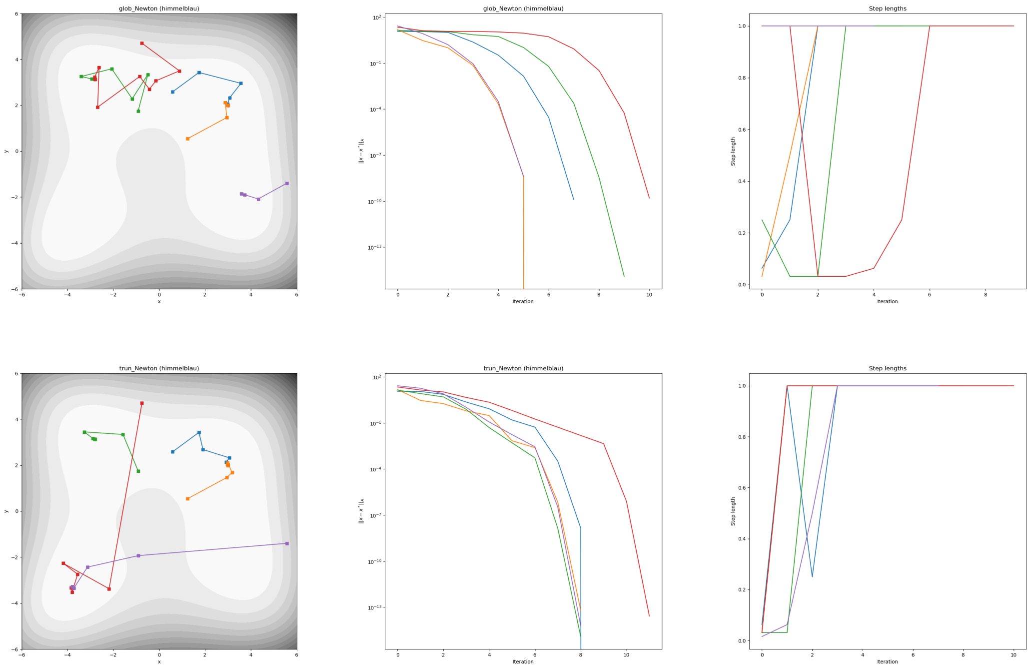

Figure 0.2: Exact Newton (top) vs truncated Newton CG (bottom) for Himmelblau function.

the TNCG compute some directions that don't pass the quality check in early iterations and the CG is actually truncated in early iterations. (6 Points)

Homework Problem 6.2 (Affine-Invariance of BFGS/DFP-updated quasi Newton) 6 Points

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable, $A \in \mathbb{R}^{n\times n}$ be invertible, $b \in \mathbb{R}^n$ and $g(y) \coloneqq f(Ay+b)$.

(i) Let the sequences $\big(x^{(k)}\big)$ and $\big(y^{(k)}\big)$ be generated by applying full quasi Newton steps as in

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - {H_f^{(k)}}^{-1} f'(x^{(k)})^\intercal && \text{from } x^{(0)} = Ay^{(0)} + b && \text{with } H_f^{(0)} \text{ s. p. d.} \\ y^{(k+1)} &= y^{(k)} - {H_g^{(k)}}^{-1} g'(y^{(k)})^\intercal && \text{from } y^{(0)} && \text{with } H_g^{(0)} = A^\intercal H_f^{(0)} A. \end{aligned}$$

Show that $x^{(k)} = Ay^{(k)} + b$ for all $k \in \mathbb{N}$, when the BFGS or the DFP update are applied to update the model Hessians.

(ii) Let the sequences $\big(x^{(k)}\big)$ and $\big(y^{(k)}\big)$ be generated by applying full inverse quasi Newton steps as in

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - B_f^{(k)} f'(x^{(k)})^\intercal && \text{from } x^{(0)} = Ay^{(0)} + b && \text{with } B_f^{(0)} \text{ s. p. d.} \\ y^{(k+1)} &= y^{(k)} - B_g^{(k)} g'(y^{(k)})^\intercal && \text{from } y^{(0)} && \text{with } B_g^{(0)} = A^{-1} B_f^{(0)} A^{-\intercal}. \end{aligned}$$

Show that $x^{(k)} = Ay^{(k)} + b$ for all $k \in \mathbb{N}$, when the inverse BFGS or the inverse DFP update are applied to update the inverse of the model Hessians.

Hint: You can save yourselves some work using the connection of the updates of the Hessians and their inverses.

Note: The restriction to unit step length scalings in this exercise is to keep the required notation slim(er). Since we know that the Armijo and the curvature condition are affine invariant as well, we don't loose invariance when applying step lengths satisfying these conditions.

Solution.

(i) The proof relies on the fact that the BFGS and DFP update formulas of the model hessians preserve the property $H_f^{(k)} = A^{-\intercal} H_g^{(k)} A^{-1}$ (the transformation behavior of the hessians of f and g in the C^2-Newton setting) for all $k \in \mathbb{N}$ – which is really the key property here.

We show the following inductively: For $k \in \mathbb{N}$ we have that $H_g^{(k)} = A^\intercal H_f^{(k)} A$ and $x^{(k)} = Ay^{(k)} + b$.

The claim is correct for $k = 0$ by assumption. In the induction step from k to $k + 1$, we can find that (due to the transformation property)

$$
\begin{aligned}
x^{(k+1)} &= x^{(k)} - {H_f^{(k)}}^{-1} f'(x^{(k)})^\intercal \\
&= Ay^{(k)} - \left(A^{-\intercal} H_g^{(k)} A^{-1}\right)^{-1} A^{-\intercal} g'(y^{(k)})^\intercal + b \\
&= A\left(y^{(k)} - {H_g^{(k)}}^{-1} g'(y^{(k)})^\intercal\right) + b \\
&= Ay^{(k+1)} + b.
\end{aligned}
$$

Now, for the updates of $H_f^{(k)}$ and $H_g^{(k)}$, we define

$$
\begin{aligned}
v_f^{(k)} &:= \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), & v_g^{(k)} &:= g'(y^{(k+1)}) - g'(y^{(k)}) = A^\intercal \nabla f(x^{(k+1)}) - A^\intercal \nabla f(x^{(k)}) = A^\intercal v_f^{(k)}, \\
r_f^{(k)} &:= x^{(k+1)} - x^{(k)} = Ar_g^{(k)}, & r_g^{(k)} &:= (y^{(k+1)}) - (y^{(k)})
\end{aligned}
$$

and otherwise reuse the notation from the lecture notes. Note that

$$
\rho_f^{(k)} = \frac{1}{(v_f^{(k)})^\intercal r_f^{(k)}} = \frac{1}{(v_g^{(k)})^\intercal A^{-1} A r_g^{(k)}} = \rho_g^{(k)}.
$$

Accordingly (red marks the term that is modified when progressing to the next line):

$$
\begin{aligned}
\Phi_{\text{BFGS}}(H_g^{(k)}, r_g^{(k)}, v_g^{(k)}) &= H_g^{(k)} - \frac{H_g^{(k)}\, r_g^{(k)}\, (r_g^{(k)})^\intercal H_g^{(k)}}{(r_g^{(k)})^\intercal H_g^{(k)}\, r_g^{(k)}} + \rho_g^{(k)}\, v_g^{(k)}\, (v_g^{(k)})^\intercal \\
&= A^\intercal H_f^{(k)} A - \frac{A^\intercal H_f^{(k)} A\, r_g^{(k)}\, (r_g^{(k)})^\intercal A^\intercal H_f^{(k)} A}{(r_g^{(k)})^\intercal A^\intercal H_f^{(k)} A\, r_g^{(k)}} + \rho_g^{(k)}\, v_g^{(k)}\, (v_g^{(k)})^\intercal \\
&= A^\intercal H_f^{(k)} A - \frac{A^\intercal H_f^{(k)} r_f^{(k)}\, (r_f^{(k)})^\intercal H_f^{(k)} A}{(r_f^{(k)})^\intercal H_f^{(k)} r_f^{(k)}} + \rho_f^{(k)}\, A^\intercal v_f^{(k)} (v_f^{(k)})^\intercal A \\
&= A^\intercal \Phi_{\text{BFGS}}(H_f^{(k)}, r_f^{(k)}, v_f^{(k)}) A.
\end{aligned}
$$

Using essentially the same computation as above but adjusting for the inverted order of the inputs $v_{f/g}^{(k)}, r_{f/g}^{(k)}$, we can immediately obtain that

$$
\Phi_{\text{BFGS}}(B_g^{(k)}, v_g^{(k)}, r_g^{(k)}) = A^{-1} \Phi_{\text{BFGS}}(B_f^{(k)}, v_f^{(k)}, r_f^{(k)}) A^{-\intercal}
$$

and therefore, using the inverse relation between the BFGS and DFP updates from (5.61) as well as that $\Phi_{\text{DFP/BFGS}}(H, s, y) = \Psi_{\text{DFP/BFGS}}(H^{-1}, y, s)^{-1}$, we have that

$$\begin{aligned}
\Phi_{\text{DFP}}(H_g^{(k)}, r_g^{(k)}, v_g^{(k)}) &= \Psi_{\text{DFP}}(B_g^{(k)}, r_g^{(k)}, v_g^{(k)})^{-1} \\
&= \Phi_{\text{BFGS}}(B_g^{(k)}, v_g^{(k)}, r_g^{(k)})^{-1} \\
&= \left(A^{-1}\Phi_{\text{BFGS}}(B_f^{(k)}, v_f^{(k)}, r_f^{(k)})A^{-\intercal}\right)^{-1} \\
&= A^\intercal \Phi_{\text{BFGS}}(B_f^{(k)}, v_f^{(k)}, r_f^{(k)})^{-1} A \\
&= A^\intercal \Psi_{\text{DFP}}(B_f^{(k)}, r_f^{(k)}, v_f^{(k)})^{-1} A \\
&= A^\intercal \Phi_{\text{DFP}}(B_f^{(k)}, r_f^{(k)}, v_f^{(k)}) A.
\end{aligned}$$

Note: We could of course have shown that property by elementary transformations, as we did for the BFGS update above.

(ii) From item (i), we know that the BFGS and DFP updated quasi Newton iterations are affine invariant. The inverse BFGS/DFP updated inverse quasi Newton iterations produce exactly the same iterates (due to the Sherman-Morrison-Woodbury formula.

(6 Points)

Homework Problem 6.3 (Inverse BFGS and DFP Updates) 6 Points

Derive the inverse BFGS and DFP update formulas

$$\Psi_{\text{BFGS}}(B, s, y) = (\text{Id} - \rho\, s\, y^\intercal)\, B\, (\text{Id} - \rho\, y\, s^\intercal) + \rho\, s\, s^\intercal, \tag{5.60}$$

$$\Psi_{\text{DFP}}(B, s, y) = B - \frac{B\, y\, y^\intercal B}{y^\intercal B\, y} + \rho\, s\, s^\intercal \tag{5.59}$$

using the Sherman-Morrison-Woodbury formula from Lemma 5.50.

Solution.

Essentially, this exercise comes down to rewriting the update structure as a rank 2 update and identifying the correct matrix roles for the Sherman-Morrison-Woodbury formula.

We begin with the inverse BFGS update (5.60). Let the difference vectors y, s be provided by the last update step corresponding to the nonsingular symmetric model Hessian H and its symmetric inverse

$B = H^{-1}$. Accordingly, the secant condition is satisfied, i. e., $Hs = y$ and $s = By$, meaning that

$$\frac{1}{\rho} = y^\mathsf{T} s = s^\mathsf{T} H s = y^\mathsf{T} B y.$$

For the BFGS update to be well defined, we need that $y^\mathsf{T} s \neq 0$, so we assume this requirement here as well.

Since H is nonsingular by assumption and in the assignment

$$\begin{aligned}
\Phi_{\text{BFGS}}(H, s, y) &= H - \frac{H\,s\,s^\mathsf{T} H}{s^\mathsf{T} H\,s} + \rho\,y\,y^\mathsf{T} \\
&= \underbrace{H}_{A \in \mathbb{R}^{n\times n}} + \underbrace{\begin{bmatrix} Hs & y \end{bmatrix}}_{U \in \mathbb{R}^{n\times 2}} \underbrace{\begin{bmatrix} -\frac{1}{s^\mathsf{T} H s} & 0 \\ 0 & \rho \end{bmatrix}}_{C \in \mathbb{R}^{2\times 2}} \underbrace{\begin{bmatrix} s^\mathsf{T} H \\ y^\mathsf{T} \end{bmatrix}}_{V = U^\mathsf{T} \in \mathbb{R}^{2\times n}}
\end{aligned}$$

the diagonal matrix $C \in \mathbb{R}^{2\times 2}$ is clearly invertible (due to the assumptions on ρ), we can apply the Sherman-Morrison-Woodbury formula

$$(A + U\,C\,V)^{-1} = A^{-1} - A^{-1}\,U\,(C^{-1} + VA^{-1}U)^{-1}\,VA^{-1}$$

to obtain that when $\Phi_{\text{BFGS}}(H, s, y)^{-1}$ is invertible (e. g. when B was s. p. d. and $y^\mathsf{T} s > 0$), then we can obtain its inverse by the update

$$
\begin{aligned}
\Phi_{\text{BFGS}}(H,s,y)^{-1} &= \left(H + \begin{bmatrix} Hs & y \end{bmatrix} \begin{bmatrix} -\frac{1}{s^\top Hs} & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} s^\top H \\ y^\top \end{bmatrix}\right)^{-1} \\
&= H^{-1} - H^{-1} \begin{bmatrix} Hs & y \end{bmatrix} \left(\begin{bmatrix} -\frac{1}{s^\top Hs} & 0 \\ 0 & \rho \end{bmatrix}^{-1} + \begin{bmatrix} s^\top H \\ y^\top \end{bmatrix} H^{-1} \begin{bmatrix} Hs & y \end{bmatrix} \right)^{-1} \begin{bmatrix} s^\top H \\ y^\top \end{bmatrix} H^{-1} \\
&= B - \begin{bmatrix} s & By \end{bmatrix} \left(\begin{bmatrix} -\frac{1}{s^\top Hs} & 0 \\ 0 & \rho \end{bmatrix}^{-1} + \begin{bmatrix} s^\top H \\ y^\top \end{bmatrix} \begin{bmatrix} s & By \end{bmatrix} \right)^{-1} \begin{bmatrix} s^\top \\ (By)^\top \end{bmatrix} \\
&= B - \begin{bmatrix} s & By \end{bmatrix} \left(\begin{bmatrix} -s^\top Hs & 0 \\ 0 & \frac{1}{\rho} \end{bmatrix} + \begin{bmatrix} s^\top Hs & y^\top s \\ y^\top s & y^\top By \end{bmatrix} \right)^{-1} \begin{bmatrix} s^\top \\ (By)^\top \end{bmatrix} \\
&= B - \begin{bmatrix} s & By \end{bmatrix} \left(\begin{bmatrix} 0 & \frac{1}{\rho} \\ \frac{1}{\rho} & \frac{1}{\rho} + y^\top By \end{bmatrix} \right)^{-1} \begin{bmatrix} s^\top \\ (By)^\top \end{bmatrix} \\
&= B + \begin{bmatrix} s & By \end{bmatrix} \begin{bmatrix} \rho + \rho^2 y^\top By & -\rho \\ -\rho & 0 \end{bmatrix} \begin{bmatrix} s^\top \\ (By)^\top \end{bmatrix} \\
&= B + \begin{bmatrix} s & By \end{bmatrix} \begin{bmatrix} \rho s^\top + \rho^2 y^\top Bys^\top - \rho(By)^\top \\ -\rho s^\top \end{bmatrix} \\
&= B + \rho ss^\top + \rho^2 sy^\top Bys^\top - \rho s(By)^\top - \rho(By)s^\top \\
&= (\mathrm{Id} - \rho\, s\, y^\top)\, B\, (\mathrm{Id} - \rho\, y\, s^\top) + \rho\, s\, s^\top \\
&= \Psi_{\text{BFGS}}(B,s,y) = \Phi_{\text{DFP}}(B,y,s)\,.
\end{aligned}
$$

To obtain the same result for the inverse DFP update, we can simply realize that we can take the inverse of that entire chain of equalities in each step, exchange B for H and vice versa, and y for s and vice versa.

(6 Points)

Homework Problem 6.4 (Two-Loop Recursion for Inverse BFGS Update) 6 Points

Show that Algorithm 5.53 in fact computes the action of the inverse BFGS updated matrix $B^{(k)}_{\text{BFGS}}$.

Solution.

We reuse the notation $\rho^{(k)} := 1/(y^{(k)})^\top s^{(k)}$ and $V^{(k)} := \mathrm{Id} - \rho^{(k)} y^{(k)} (s^{(k)})^\top$ from the lecture notes and note that the inverse BFGS update can then be written as

$$B^{(k+1)}_{\text{BFGS}} = (V^{(k)})^\top B^{(k)}_{\text{BFGS}} V^{(k)} + \rho^{(k)} s^{(k)} (s^{(k)})^\top.$$

For a fixed $k \in \mathbb{N}$ and an input vector $r \in \mathbb{R}^n$, the two-loop recursion in Algorithm 5.53 computes the quantities $q^{(i)}$ by setting (in the following the loop index i facilitates the move from the counter i to $i-1$ or $i+1$, respectively):

$$\begin{aligned}
i = k \quad &: \quad q^{(k)} && := r \\
i = k-1 : \quad & \quad q^{(k-1)} && := V^{(k-1)} q^{(k)} \\
&\vdots \\
i \quad &: \quad q^{(i)} && := V^{(i)} q^{(i+1)} \\
&\vdots \\
i = 0 \quad &: \quad q^{(0)} && := V^{(0)} q^{(1)}
\end{aligned}$$

in the backward loop storing the $\alpha^{(i)}$ but always rewriting the $q^{(i)}$ to the storage of r. In the forward loop, the values $z^{(i)}$ are computed as

$$\begin{aligned}
i = 0 : \quad & z^{(0)} && := B^{(0)}_{\text{BFGS}} q^{(0)} \\
i = 1 : \quad & z^{(1)} && := z^{(0)} + (\alpha^{(0)} - \beta^{(0)}) s^{(0)} \\
&\vdots \\
i+1 \quad : \quad & z^{(i+1)} && := z^{(i)} + (\alpha^{(i)} - \beta^{(i)}) s^{(i)} \\
&\vdots \\
i = k : \quad & z^{(k)} && := z^{(k)} + (\alpha^{(k)} - \beta^{(k)}) s^{(k)}.
\end{aligned}$$

Where we can inductively show that $z^{(i)} = B^{(i)}_{\text{BFGS}} q^{(i)}$, because

$$\begin{aligned}
z^{(i+1)} &:= z^{(i)} + (\alpha^{(i)} - \beta^{(i)}) s^{(i)} \\
&= z^{(i)} - \underbrace{\beta^{(i)}}_{\rho^{(i)} (y^{(i)})^\intercal z^{(i)}} s^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal z^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal B^{(i)}_{\text{BFGS}} q^{(i)} + \alpha^{(i)} s^{(i)} \\
&= (V^{(i)})^\intercal B^{(i)}_{\text{BFGS}} V^{(i)} q^{(i+1)} + \rho^{(i)} (s^{(i)})^\intercal q^{(i+1)} s^{(i)} \\
&= B^{(i+1)}_{\text{BFGS}} q^{(i+1)}
\end{aligned}$$

so after k iterations, we end up with $B^{(k)}_{\text{BFGS}} q^{(k)} = B^{(k)}_{\text{BFGS}} r$ as expected. (6 Points)

Please submit your solutions as a single pdf and an archive of programs via moodle.