Response to Second Referee on the Paper: A modified nearly exact method for solving low-rank trust region subproblem

First, we would like to thank you for taking the time and effort to review and provide comments to our paper. We appreciate your willingness to help to make our paper a better, more worthwhile document.

Below we answer the two comments you have given in your report.

1. "However, we will have no hard-case if the Hessian is positive definite (particularly when the matrix is generated by L-BFGS).": Clearly, when the Hessian of the LRTR subproblem is positive definite, its solution does not fall into the hard case. When line search methods with specific stepsize rules (e.g., Wolfe or strong Wolfe rule) are used, then the Hessian generated by the L-BFGS method is guaranteed to be positive definite. However, when the iterates are found using the trust region strategy, the Hessian generated by the L-BFGS method is not necessarily positive definite. In fact, in our implementation the hard case occurs quite frequently for some NLP instances.

2. "Related earlier work is as follows. The standard trust region subproblem is actually equivalent to a low dimensional trust region subproblem if the Hessian is generated by L-BFGS method (for example, see Wang, Wen and Yuan (2004); Yuan (2004)).": First, our present work was done independently from the works by Wang, Wen and Yuan (2004) and Yuan (2004). A preliminary draft of our paper was completed in 2003, and a talk based on it was given in the Large-Scale Nonlinear Programming Session, 2003 INFORMS Annual Meeting, Atlanta, USA, October 19-22, 2003. We were not aware of the above two references and we are certainly grateful to the referee for bringing them to our attention. We have cited them in our paper. Second, the method proposed in Wang, Wen and Yuan (2004) and Yuan (2004) is generally different from our limited memory quasi-Newton TR method based on solving a sequence of **full dimensional** LRTR subproblems since the TR subproblem of one method may not be converted into an equivalent one of the other method (see our discussions below). Third, **full dimensional** LRTR subproblems naturally appear when using limited memory quasi-Newton TR method for solving constrained NLP. For example, when solving NLP with simple bound constraints, the Dikin ellipsoid type of trust regions are extensively used (see the discussions on pp. 3-4 and pp. 18 of our paper).

First, we argue that the LRTR subproblem studied in our paper generally cannot be converted into an equivalent TR subproblem discussed in Wang, Wen and Yuan (2004). Indeed, consider a

LRTR subproblem (see also equations (1)-(4) of our paper)

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}p^T H p + g^T p \\ \text{s.t.} & \|p\|_M \le \Delta \end{array} \tag{1}$$

where

$$H = D + VEV^T, \tag{2}$$
$$M = \tilde{D} + \tilde{V}\tilde{E}\tilde{V}^T \succ 0, \tag{3}$$

and D, $\tilde{D}$ and $\tilde{E}$ are positive diagonal matrices, V and $\tilde{V}$ have few number of columns (say less than 10), and E is a diagonal matrix.

Case 1: Assume first that $D^{-1}\tilde{D} \in \{\gamma I : \gamma > 0\}$. In this case, we claim that there exists an optimal solution p^* of LRTR subproblem (1) such that $p^* \in S$, where

$$S \equiv \text{Range}(A), \quad A \equiv D^{-1}[-g\ V\ \tilde{V}]. \tag{4}$$

(Its proof will be given in the next paragraph.) Hence, using a similar argument as in Wang, Wen and Yuan (2004), we can easily convert the LRTR subproblem (1) to an equivalent TR subproblem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}z^T A^T H A z + (A^T g)^T z \\ \text{s.t.} & \|z\|_{\hat{M}} \le \Delta \end{array} \tag{5}$$

where $\hat{M} = A^T M A$. However, we shall notice that the TR subproblem (5) is generally **not** a low dimensional TR subproblem as defined in equation (2.13) of Wang, Wen and Yuan (2004) since $\hat{M} \neq I$. Thus, for the special case, even a standard TR subproblem with the Hessian H generated by L-BFGS method (i.e., $M = I$ and $D = \gamma I$ for some $\gamma > 0$) is **not** equivalent to a low dimensional TR subproblem as discussed in Wang, Wen and Yuan (2004).

We next sketch the proof of the above claim. Let

$$\lambda_1 \equiv \lambda_{\min}(M^{-1/2} H M^{-1/2}).$$

(see also equation (13) of our paper.) Let $H(\lambda) \equiv H + \lambda M$. First, we will show that for any $\lambda > -\lambda_1$ (or equivalently, $H(\lambda) \succ 0$),

$$H(\lambda)^{-1} w \in S, \quad \forall w \in \text{Range}([g\ V\ \tilde{V}]). \tag{6}$$

Indeed, using (2), (3), Sherman-Morrison-Woodbury (SMW) formula, and the assumption that $D^{-1}\tilde{D} = \gamma I$ for some $\gamma > 0$, we observe that, for $0 \neq \lambda \in (-\lambda_1, \infty)$ and $w \in \Re^n$,

$$\begin{aligned} H(\lambda)^{-1} w &= (D + VEV^T + \lambda\tilde{D} + \lambda\tilde{V}\tilde{E}\tilde{V}^T)^{-1} w \\ &= \left(D + \lambda\tilde{D} + [V\ \tilde{V}] \begin{bmatrix} E & 0 \\ 0 & \lambda\tilde{E} \end{bmatrix} \begin{bmatrix} V^T \\ \tilde{V}^T \end{bmatrix} \right)^{-1} w \\ &= \left((D + \lambda\tilde{D})^{-1} - (D + \lambda\tilde{D})^{-1} [V\ \tilde{V}] Q^{-1} \begin{bmatrix} V^T \\ \tilde{V}^T \end{bmatrix} (D + \lambda\tilde{D})^{-1} \right) w \\ &= (D + \lambda\tilde{D})^{-1} \left(w - [V\ \tilde{V}] Q^{-1} \begin{bmatrix} V^T \\ \tilde{V}^T \end{bmatrix} (D + \lambda\tilde{D})^{-1} w \right) \\ &= (1 + \gamma\lambda)^{-1} D^{-1} \left(w - [V\ \tilde{V}] Q^{-1} \begin{bmatrix} V^T \\ \tilde{V}^T \end{bmatrix} (D + \lambda\tilde{D})^{-1} w \right), \end{aligned}$$

where

$$Q = \left[\begin{array}{cc} E & 0 \\ 0 & \lambda\tilde{E} \end{array} \right]^{-1} + \left[\begin{array}{c} V^T \\ \tilde{V}^T \end{array} \right] (D+\lambda\tilde{D})^{-1}[V \ \tilde{V}].$$

The last equality above together with (4) immediately implies that (6) holds for $0 \neq \lambda \in (-\lambda_1, \infty)$. Similarly, (6) also holds for $\lambda = 0$ if $0 \in (-\lambda_1, \infty)$. From the discussion on pp. 6 of our paper, we know that three cases may occur for LRTR subproblem (1), that is, the "easy", "interior convergence", and "hard" cases. We know that, in the first two cases, the optimal solution p^* of LRTR subproblem (1) is unique and $p^* = -H(\lambda)^{-1}g$ for some $\lambda > -\lambda_1$, and hence $p^* \in S$ due to (6). In the "hard" case, we also see from the discussion on pp. 6 of our paper that there exists an optimal solution p^* of LRTR subproblem (1) such that $p^* = p_{\rm crt} + \alpha^M u$ for some $\alpha^M \in \Re$ and $0 \neq u \in {\rm Ker}\,(H(-\lambda_1))$, where $p_{\rm crt} \equiv -H(-\lambda_1)^{\dagger}g$, and the superscript † denotes the Moore-Penrose generalized inverse. Using (6) and the fact that the subspace S is closed, we easily see that

$$p_{\rm crt} = -\lim_{\lambda \downarrow -\lambda_1} H(\lambda)^{-1}g \in S. \tag{7}$$

Recall from (2) that $H = D + VEV^T$, where $D \succ 0$ and E is diagonal and nonsingular. We can partition E (after performing a symmetric permutation of its rows and columns) as $E = {\rm Diag}\,(E_1, -E_2)$, where both E_1 and E_2 are positive diagonal matrices. Accordingly, we partition V as $V = (V_1, V_2)$, and hence

$$VEV^T = V_1E_1V_1^T - V_2E_2V_2^T$$

(see also Subsection 3.2 and equation (28) of our paper). For any $\lambda > -\lambda_1$, we let

$$u_\lambda = H(\lambda)^{-1}v/\|H(\lambda)^{-1}v\|_M, \tag{8}$$

where $v = V_2r$ and r is a unit eigenvector of $V_2^TH(\lambda)^{-1}V_2$ corresponding to its maximum eigenvalue (see also Theorem 3.3 of our paper). Using the fact that $v = V_2r \in {\rm Range}([g \ V \ \tilde{V}])$, we easily see from (6) that $u_\lambda \in S$ for any $\lambda > -\lambda_1$. Let u be any accumulation point of u_λ as $\lambda \downarrow -\lambda_1$, and hence $u \in S$. Further, using Theorem 3.3 of our paper, we have $u^TH(-\lambda_1)u = 0$, i.e., $u \in {\rm Ker}(H(-\lambda_1))$. Therefore, $u \in {\rm Ker}(H(-\lambda_1)) \cap S$, and $u \neq 0$ since $\|u\|_M = 1$. This together with (7) and the relation $p^* = p_{\rm crt} + \alpha^M u$ implies that $p^* \in S$. Hence, the above claim holds.

Case 2: $D^{-1}\tilde{D} \notin \{\gamma I | \gamma > 0\}$**.** In this case, we will see from the following observation that it is hard to identify a fixed subspace where an optimal solution of LRTR subproblem (1) lies unless the optimal λ^* is known. Using (2), (3), and SMW formula, we see that for any $\lambda > -\lambda_1$,

$$-H(\lambda)^{-1}g \in S_\lambda,$$

where $S_\lambda \equiv (I + \lambda D^{-1}\tilde{D})^{-1}S$, and S is defined in (4). We shall notice that the subspace S_λ varies as λ changes over $(-\lambda_1, \infty)$. Therefore, in this case, the LRTR subproblem (1) certainly cannot be converted into an equivalent low dimensional TR subproblem studied in Wang, Wen and Yuan (2004).

Second, the TR subproblem discussed in Wang, Wen and Yuan (2004) may not be converted into an equivalent **full dimensional** LRTR subproblem studied in our paper (see our discussions below).

For convenience, we rewrite the TR subproblem discussed in Wang, Wen and Yuan (2004) in the form:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}p^T H p + g^T p \\ \text{s.t.} \quad & p = Bz, \\ & \|z\| \le \Delta, \end{aligned} \tag{9}$$

where H is as defined in (2) with $D = \gamma I$ for some $\gamma > 0$, and $B = [-g \ V]$. For convenience, we assume that B has a full column rank. Under this assumption, for any feasible solution (p, z) of (9), we have

$$p = Bz \ \Rightarrow \ z = (B^T B)^{-1} B^T p,$$

and hence $\|(B^T B)^{-1} B^T p\| \le \Delta$ since $\|z\| \le \Delta$. Let $\tilde{M} = B(B^T B)^{-2} B^T$. Then, TR subproblem (9) is equivalent to the TR subproblem:

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2}p^T H p + g^T p \\ \text{s.t.} \quad & \|p\|_{\tilde{M}} \le \Delta. \end{aligned} \tag{10}$$

We shall notice that $\tilde{M} \succeq 0$, but it may not be positive definite. Thus, the TR subproblem (10) is not a special case of the LRTR subproblem studied in our paper.

Finally, the main contribution of this paper is to show that Moré and Sorensen's algorithm can be efficiently modified to solve large scale LRTR subproblems. The most interesting contribution of this paper in this respect is the handling of the hard case. The paper also demonstrates the efficiency of the proposed method when used to solve large scale NLP problems. We have not addressed the issue of global convergence of the overall trust region method since we believe this issue would easily follow from well-known arguments used in the convergence analysis of trust region methods.

Again, thank you for your insightful comments.