A Highly Scalable Parallel Algorithm for Isotropic Total Variation Models (Supplemental Material)

Jie Wang JIE.WANG.USTC@ASU.EDU

Arizona State University, Tempe, AZ 85287 USA

Qingyang Li QINGYANG.LI@ASU.EDU

Arizona State University, Tempe, AZ 85287 USA

Sen Yang SENYANG@ASU.EDU

Arizona State University, Tempe, AZ 85287 USA

Wei Fan DAVID.FANWEI@HUAWEI.COM

Huawei Noahs Ark Lab, Hong Kong, China

Peter Wonka PWONKA@GMAIL.COM

Arizona State University, Tempe, AZ 85287 USA

Jieping Ye JIEPING.YE@ASU.EDU

Arizona State University, Tempe, AZ 85287 USA

In this supplement, we present all the necessary details we mentioned in the main text.

1. Algorithm of ADMM

Denote K the number of iterations required to meet the stopping criterion. The update rule in each iteration is given by Algorithm 1. For more details, please refer to **?**.

Algorithm 1 ADMM

Input: $\mathbf{z}_0, \theta_0, \gamma > 0$

for $t = 0$ **to** T **do**

$$\mathbf{x}^{t+1} := \operatorname*{argmin}_{\mathbf{x}} L_\gamma(\mathbf{x}, \mathbf{z}^t; \theta^t) \tag{1}$$

$$\mathbf{z}^{t+1} := \operatorname*{argmin}_{\mathbf{z}} L_\gamma(\mathbf{x}^{t+1}, \mathbf{z}; \theta^t) \tag{2}$$

$$\theta^{t+1} := \theta^t + (\mathbf{x}^{t+1} - \mathbf{z}^{t+1}) \tag{3}$$

end for

Return: $\mathbf{x}^T, \mathbf{z}^T$

---

2. Convergence Analysis

The standard form of ADMM is

$$\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + g(\mathbf{z}) \qquad \text{s.t. } \mathbf{x} - \mathbf{z} = 0 \tag{4}$$

Algorithm FAD finds the solution of the the following constrained convex optimization problem:

$$\begin{aligned} \min_{Z;X_1,X_2,X_3} \quad & \frac{1}{2}\|Z - Y\|_F^2 + \lambda \sum_{k=1}^{3} \|X_k\|_{TV_k} \\ \text{s.t.} \quad & X_k = Z,\ k = 1, 2, 3. \end{aligned} \tag{5}$$

The convergence properties of ADMM to solve the standard form in (4) have been extensively explored by **???**. Therefore, to establish the convergence properties of Algorithm FAD, we only need to reformulate problem (5) as (4) and check if the resulting formulation satisfies the conditions required for convergence.

Let us denote $\mathbf{z} := (Z^T, Z^T, Z^T)^T$, $\mathbf{x} := (X_1^T, X_2^T, X_3^T)^T$, and $T_1 := (I_m, 0, 0)$, $T_2 := (0, I_m, 0)$, $T_3 := (0, 0, I_m)$, where $I_m \in \Re^{m\times m}$ is the identity matrix, and $I \in \Re^{3m\times m} = (I_m, I_m, I_m)^T$. Then problem (5) can be rewritten as:

$$\min_{\mathbf{x},\mathbf{z}} f(\mathbf{x}) + g(\mathbf{z}) \qquad \text{s.t. } \mathbf{x} - \mathbf{z} = 0, \tag{6}$$

where

$$f(\mathbf{x}) = \lambda \sum_{k=1}^{3} \|T_k \mathbf{z}\|_{TV_k},$$

$$g(\mathbf{z}) = \frac{1}{6}\|\mathbf{z} - IY\|_F^2.$$

Clearly, problem in (6) is exactly the same as (4). Since f and g are proper closed convex functions, the convergence properties of Algorithm FAD can be readily established by **???**. The convergence rate of Algorithm FAD can be shown as $O(1/k)$ by following the procedure in **?**.

3. Dual Formulation

Recall that we need to solve

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + \rho\sqrt{(u_1 - u_2)^2 + (u_3 - u_2)^2}, \tag{7}$$

where $\rho = \frac{\lambda}{\gamma}$.

3.1. Deviation of the Dual Formulation

Without loss of generality, assume $\mathbf{w} \in \Re^{2n+1}$, we solve the following optimization problem:

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + \rho \sum_{k=1}^{n} \sqrt{(u_{2k-1} - u_{2k})^2 + (u_{2k+1} - u_{2k})^2} \tag{8}$$

Clearly, problem (8) reduces to problem (7) when $n = 1$. Let $\mathbf{z} \in \Re^{2n}$ be defined as:

$$z_i = u_{i+1} - u_i,\ i = 1, 2, \ldots, 2n.$$

Then problem (8) is equivalent to the following constrained optimization problem:

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + \rho \sum_{k=1}^{n} \sqrt{z_{2k-1}^2 + z_{2k}^2} \qquad \text{s.t. } z_i = u_{i+1} - u_i,\ i = 1, 2, \ldots, 2n. \tag{9}$$

Let $G \in \Re^{2n\times(2n+1)}$ be defined as

$$g_{i,j} = \begin{cases} 1 & \text{if } j = i+1 \\ -1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

we have $\mathbf{z} = G\mathbf{u}$.

By introducing the scaled dual variable $\rho\mathbf{s} \in \Re^{2n}$, the Lagrangian of problem (8) can be written as:

$$L(\mathbf{u},\mathbf{z};\mathbf{s}) = \frac{1}{2}\|\mathbf{u}-\mathbf{w}\|_2^2 + \rho\sum_{k=1}^{n}\sqrt{z_{2k-1}^2 + z_{2k}^2} + \rho\langle \mathbf{s}, G\mathbf{u}-\mathbf{z}\rangle \tag{11}$$

and the dual problem is in fact

$$g(\mathbf{s}) = \min_{\mathbf{u},\mathbf{z}} L(\mathbf{u},\mathbf{z};\mathbf{s}). \tag{12}$$

Let

$$f_1(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}-\mathbf{w}\|_2^2 + \rho\langle \mathbf{s}, G\mathbf{u}\rangle$$

and

$$f_2(\mathbf{z}) = \rho\sum_{k=1}^{n}\sqrt{z_{2k-1}^2 + z_{2k}^2} - \rho\langle \mathbf{s}, \mathbf{z}\rangle$$

We can see that

$$L(\mathbf{u},\mathbf{z};\mathbf{s}) = f_1(\mathbf{u}) + f_2(\mathbf{z})$$

and

$$g(\mathbf{s}) = \min_{\mathbf{u}} f_1(\mathbf{u}) + \min_{\mathbf{z}} f_2(\mathbf{z}) \tag{13}$$

Let us first solve the optimization problem $\min_{\mathbf{u}} f_1(\mathbf{u})$. Since $\nabla f_1(\mathbf{u}) = \mathbf{u} - \mathbf{w} + \rho G^T\mathbf{s}$, by setting $\nabla f_1(\mathbf{u}) = 0$, we have

$$\mathbf{u}^* = \operatorname*{argmin}_{\mathbf{u}} f_1(\mathbf{u}) = \mathbf{w} - \rho G^T\mathbf{s} \tag{14}$$

and

$$f_1(\mathbf{u}^*) = \min_{\mathbf{u}} f_1(\mathbf{u}) = -\frac{\rho^2}{2}\|G^T\mathbf{s}\|_2^2 + \rho\langle G^T\mathbf{s}, \mathbf{w}\rangle = -\frac{1}{2}\|\mathbf{w}-\rho G^T\mathbf{s}\|_2^2 + \frac{1}{2}\|\mathbf{w}\|_2^2$$

To solve the second optimization problem $\min_{\mathbf{z}} f_2(\mathbf{z})$, let us make some conventions. For an arbitrary vector $\mathbf{v}$, let $[\mathbf{v}]_i^j = (v_i, v_{i+1}\ldots, v_j)^T$. Then $f_2(\mathbf{z})$ can be written as:

$$f_2(\mathbf{z}) = \rho\sum_{k=1}^{n}\|[\mathbf{z}]_{2k-1}^{2k}\|_2 - \rho\langle \mathbf{s}, \mathbf{z}\rangle \tag{15}$$

and thus

$$[\nabla f_2(\mathbf{z})]_{2k-1}^{2k} \in \begin{cases} \rho\left\{\frac{[\mathbf{z}]_{\mathbf{2k-1}}^{\mathbf{2k}}}{\|[\mathbf{z}]_{\mathbf{2k-1}}^{\mathbf{2k}}\|_{\mathbf{2}}} - [\mathbf{s}]_{2k-1}^{2k}\right\}, \text{ if } [\mathbf{z}]_{\mathbf{2k-1}}^{\mathbf{2k}} \neq \mathbf{0} \\ \rho\left\{\mathbf{v}_k - [\mathbf{s}]_{2k-1}^{2k}\right\}, \|\mathbf{v}_k\|_2 \leq 1, \text{ if } [\mathbf{z}]_{\mathbf{2k-1}}^{\mathbf{2k}} = \mathbf{0} \end{cases} \tag{16}$$

For each $k \in \{1,\ldots,n\}$, by setting $[\nabla f_2(\mathbf{z})]_{2k-1}^{2k} = 0$, it follows that

$$\|[\mathbf{s}]_{2k-1}^{2k}\|_2 \leq 1 \tag{17}$$

and

$$f_2(\mathbf{z}^*) = \min_{\mathbf{z}} f_2(\mathbf{z}) = 0,$$

where $\mathbf{z}^* = \operatorname{argmin}_{\mathbf{z}} f_2(\mathbf{z})$.

Therefore, the dual function $g(\mathbf{s})$ is found as:

$$g(\mathbf{s}) = -\frac{1}{2}\|\mathbf{w}-\rho G^T\mathbf{s}\|_2^2 + \frac{1}{2}\|\mathbf{w}\|_2^2 \tag{18}$$

All together, the dual problem of problem (8) is equivalent to the following optimization problem:

$$\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{w}-\rho G^T\mathbf{s}\|_2^2 \tag{19}$$
$$\text{s.t. } \|[\mathbf{s}]_{2k-1}^{2k}\|_2 \leq 1,\ k = 1,2,\ldots,n.$$

3.2. KKT Conditions

The KKT conditions can be directly obtained from the last section. Assume $\mathbf{u}^*$, $\mathbf{z}^*$ and $\mathbf{s}^*$ are the optimal solutions to the problems (8) and (19) respectively, we summarize the KKT conditions as follows:

$$\mathbf{w} = \mathbf{u}^* + \rho G^T\mathbf{s}^* \tag{20}$$

$$[\mathbf{s}^*]_{2k-1}^{2k} \in \begin{cases} \frac{[\mathbf{z}^*]_{2k-1}^{2k}}{\|[\mathbf{z}^*]_{2k-1}^{2k}\|_2}, \text{if } [\mathbf{z}^*]_{2k-1}^{2k} \neq 0 \\ \mathbf{v}_k, \|\mathbf{v}_k\|_2 \leq 1, \text{if } [\mathbf{z}^*]_{2k-1}^{2k} = 0 \end{cases} \tag{21}$$

Because $\mathbf{z}^* = G\mathbf{u}^*$, condition (21) can be written more explicitly as:

$$[\mathbf{s}^*]_{2k-1}^{2k} \in \begin{cases} \frac{(u_{2k}-u_{2k-1}, u_{2k+1}-u_{2k})^T}{\|(u_{2k}-u_{2k-1}, u_{2k+1}-u_{2k})^T\|_2}, \text{if } (u_{2k}-u_{2k-1}, u_{2k+1}-u_{2k})^T \neq 0 \\ \mathbf{v}_k, \|\mathbf{v}_k\|_2 \leq 1, \text{if } (u_{2k}-u_{2k-1}, u_{2k+1}-u_{2k})^T = 0 \end{cases} \tag{22}$$

4. Proposition 1

Proposition 1. *Given an image with $m \times n$ pixels, if we divide the image domain into three non-overlapping subset*

$$\mathcal{D}_k := \{(i,j) \in \mathcal{D} : \mod (j-i, 3) = k-1\}, \tag{23}$$

where $k = 1, 2, 3$, and we define

$$\|X\|_{TV}^{(k)} := \sum_{(i,j)\in\mathcal{D}_k} \|D_{i,j}\|_2, \quad k = 1, 2, 3,$$

then for each $(i,j) \in \mathcal{D}_k$, $\|D_{i,j}\|_2$ is separable from every other $\|D_{i',j'}\|_2$, where $(i', j') \in \mathcal{D}_k \setminus (i,j)$.

Proof. To simplify the proof, let us extend $\mathcal{D}$ to an infinite set, i.e.,

$$\mathcal{D} = \{(i,j) : \quad i = -\infty, \ldots, -1, 0, 1, \ldots, \infty, \\ j = -\infty, \ldots, -1, 0, 1, \ldots, \infty.\}$$

If the conclusion of Proposition 1 is true for the extended $\mathcal{D}$, then it also holds when $\mathcal{D}$ is a finite set. Therefore let us prove Proposition 1 for the former case.

Without loss of generality, let us consider $\mathcal{D}_1$. Every term $\|D_{i,j}\|_2$ in $\|X\|_{TV}^{(1)}$ involves three variables, $x_{i,j}$, $x_{i+1,j}$ and $x_{i,j+1}$ since the pixels (i,j), $(i+1,j)$ and $(i,j+1)$ all belong to $\mathcal{D}_1$ (we extend $\mathcal{D}$ to an infinite set).

Consider variable $x_{i,j}$. It is only appears in $\|D_{i-1,j}\|_2$ and $\|D_{i,j-1}\|$. Since $(i,j) \in \mathcal{D}_1$, we have $\mod (j-i, 3) = 0$. Therefore it follows that

$$\mod (j-(i-1), 3) = 1 \text{ and } \mod ((j-1)-i, 3) = 2.$$

According to the definition of $\mathcal{D}_2$ and $\mathcal{D}_3$, we can see that $(i-1,j) \in \mathcal{D}_2$ and $(i,j-1) \in \mathcal{D}_3$. In other words, except $\|D_{i,j}\|_2$, the variable $x_{i,j}$ does not appear in any other term $\|D_{i',j'}\|_2$ where $(i',j') \in \mathcal{D}_1$. For the same reason, the variables $x_{i-1,j}$ and $x_{i,j-1}$ only appears in $\|D_{i,j}\|_2$ among all $\|D_{i'',j''}\|_2$ where $(i'',j'') \in \mathcal{D}_1$.

Therefore $\|D_{i,j}\|_2$ is separable from all the other terms in $\|X\|_{TV}^{(1)}$, which completes the proof. □

5. Proof of Theorem 1

Proof. Denote the objective function of as $f(\mathbf{s})$. Suppose ρ is large enough such that $\mathbf{t}^*(\rho) \in \operatorname{relint} \widetilde{\mathcal{B}}$ and thus $\mathbf{s}^*(\rho) \in \mathcal{B}^\circ$. Then we have

$$\nabla f(\mathbf{s}^*) = -\rho G(\mathbf{w} - \rho G^T \mathbf{s}^*) = 0, \tag{24}$$

which is equivalent to

$$(GG^T)^{-1} G\mathbf{w} = \rho \mathbf{s}^*. \tag{25}$$

Therefore, we have

$$\|(GG^T)^{-1} G\mathbf{w}\|_2 = \rho \|\mathbf{s}^*\|_2.$$

It follows that

$$\rho > \|(GG^T)^{-1} G\mathbf{w}\|_2 \Rightarrow \|\mathbf{s}^*\|_2 < 1.$$

Consider the KKT condition $\mathbf{u}^* = \mathbf{w} - \rho G^T \mathbf{s}^*$ together with Eq. (25), $\mathbf{u}^*$ is:

$$\mathbf{u}^* = \mathbf{w} - G^T (GG^T)^{-1} G\mathbf{w},$$

which is the result in the statement. Note that $\mathbf{u}^*$ is independent of ρ.

Recall that $\mathbf{s}^*(\rho)$ varies continuously with ρ. By the KKT condition $\mathbf{u}^* = \mathbf{w} - \rho G^T \mathbf{s}^*$, $\mathbf{u}^*$ is also continuous with respect to ρ. Thus

$$\mathbf{u}^*(\rho_{max}) = \lim_{\rho \downarrow \rho_{max}} \mathbf{u}^*(\rho) = \mathbf{w} - G^T(GG^T)^{-1}G\mathbf{w},$$

which completes the proof. □

6. Proof of Theorem 2

Proof. Let $\mathbf{s}^*$ be the optimal solution to the dual problem. The KKT conditions are:

$$-\rho G(\mathbf{w} - \rho G^T\mathbf{s}^*) + \alpha\mathbf{s}^* = 0 \tag{26}$$

$$\alpha((s_1^*)^2 + (s_2^*)^2 - 1) = 0 \tag{27}$$

where $\alpha > 0$ is the Lagrange multiplier and Eq. (27) is the complementary condition.

By rearranging the terms, Eq. (26) becomes

$$\rho(\rho^2 GG^T + \alpha I)^{-1}G\mathbf{w} = \mathbf{s}^*.$$

It follows that

$$\begin{aligned} &\|\rho(\rho^2 GG^T + \alpha I)^{-1}G\mathbf{w}\|_2^2 = \|\mathbf{s}^*\|_2^2 \\ \Rightarrow &\left\|\rho\frac{1}{\rho^2}\left(U\Sigma^2U^T + \frac{\alpha}{\rho^2}UU^T\right)^{-1}G\mathbf{w}\right\|_2^2 = \|\mathbf{s}^*\|_2^2 \\ \Rightarrow &\left\|U\left(\Sigma^2 + \frac{\alpha}{\rho^2}I\right)^{-1}\Sigma V^T\mathbf{w}\right\|_2^2 = \rho^2\|\mathbf{s}^*\|_2^2 \\ \Rightarrow &\left\|\left(\Sigma^2 + \frac{\alpha}{\rho^2}I\right)^{-1}\widetilde{\mathbf{w}}\right\|_2^2 = \rho^2\|\mathbf{s}^*\|_2^2 \end{aligned} \tag{28}$$

where

$$\Sigma^2 = \Sigma\Sigma^T = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \widetilde{\mathbf{w}} = \Sigma V^T\mathbf{w}.$$

Eq. (28) is equivalent to

$$\left(\frac{\rho\widetilde{w}_1}{\rho^2\sigma_1^2 + \alpha}\right)^2 + \left(\frac{\rho\widetilde{w}_2}{\rho^2\sigma_2^2 + \alpha}\right)^2 = \|\mathbf{s}^*\|_2^2 \tag{29}$$

Let $\mathbf{v} \in \Re^2$, where

$$v_1 = \frac{\rho\widetilde{w}_1}{\rho^2\sigma_1^2 + t} \text{ and } v_2 = \frac{\rho\widetilde{w}_2}{\rho^2\sigma_2^2 + t}.$$

Clearly, $\mathbf{v}(t)$ is a plane curve parameterized by $t \geq 0$ and it is easy to see that

$$\lim_{t\to\infty}\|\mathbf{v}(t)\|_2^2 = 0 \text{ and } \frac{d\|\mathbf{v}(t)\|_2^2}{dt} < 0, \ \forall t \geq 0.$$

Therefore, by Eq. (29), we can see that:

1) If $\|\mathbf{v}(0)\|_2^2 < 1$, then $\|\mathbf{s}^*\|_2^2 = \|\mathbf{v}(\alpha)\|_2^2 < 1$. The complementary condition Eq. (27) indicates that $\alpha = 0$. Then Eq. (26) reduces to Eq. (25), which has been solved in Theorem 1.

It is worthwhile to note that

$$\begin{aligned} \|\mathbf{v}(0)\|_2^2 &= \|(\rho\Sigma^2)^{-1}\widetilde{\mathbf{w}}\|_2^2 = \frac{1}{\rho^2}\|U(\Sigma V^TV\Sigma^T)^{-1}U^TU\widetilde{\mathbf{w}}\|_2^2 \\ &= \frac{1}{\rho^2}\|(U\Sigma V^TV\Sigma^TU^T)^{-1}U\widetilde{\mathbf{w}}\|_2^2 = \frac{1}{\rho^2}\|(GG^T)^{-1}\mathbf{w}\|_2^2 \end{aligned}$$

Therefore, from $\|\mathbf{v}(0)\|_2^2 < 1$, we can derive that $\rho > \|(GG^T)^{-1}\mathbf{w}\|_2 = \rho_{max}$ which is consistent with the results in Theorem 1.

2) If $\|\mathbf{v}(0)\|_2^2 > 1$, we must have $\alpha > 0$ since $\|\mathbf{v}(\alpha)\|_2 = \|\mathbf{s}^*\|_2 \leq 1$ [recall the constraint of problem (26) in the main text]. The complementary condition implies that $\|\mathbf{s}^*\|_2 = 1$. Clearly, the curve $\mathbf{v}(t)$, $t \geq 0$ must intersect with the unit circle at one unique point. Therefore Eq. (29) must have one unique positive root. Furthermore, from $\|\mathbf{v}(0)\|_2^2 > 1$, we can conclude that $\rho < \|(GG^T)^{-1}\mathbf{w}\|_2 = \rho_{max}$.

From the above, we know $\rho < \rho_{max}$ implies that $\|\mathbf{s}^*\|_2 = 1$. Sorting the terms in Eq. (29) results in Eq. (30), which completes the proof. □

In the proof of Theorem 2, we omit the case $\|\mathbf{v}(0)\|_2 = 1$. By the same technique, we know $\|\mathbf{v}(0)\| = 1$ implies $\rho = \rho_{max}$, which has been considered in Theorem 1. Actually, via the techniques in Theorem 2, the case $\rho = \rho_{max}$ is equivalent to that Eq. (30) has a unique root at 0.

7. Discussion of Root Finding Technique

The quartic function we need to solve is solves the following quartic function:

$$\alpha^4 + c_3\alpha^3 + c_2\alpha^2 + c_1\alpha + c_0 = 0 \tag{30}$$

where

$$\begin{aligned} c_0 &= \rho^8\sigma_1^4\sigma_2^4 - \rho^6(\widetilde{w}_1^2\sigma_2^4 + \widetilde{w}_2^2\sigma_1^4), \\ c_1 &= 2\rho^6(\sigma_1^2\sigma_2^4 + \sigma_1^4\sigma_2^2) - 2\rho^4(\widetilde{w}_1^2\sigma_2^2 + \widetilde{w}_2^2\sigma_1^2), \\ c_2 &= \rho^4(\sigma_1^4 + \sigma_2^4 + 4\sigma_1^2\sigma_2^2) - \rho^2(\widetilde{w}_1^2 + \widetilde{w}_2^2), \\ c_3 &= 2\rho^2(\sigma_1^2 + \sigma_2^2). \end{aligned}$$

In Theorem 2, we have already shown that Eq. (30) has a unique positive root when $\rho < \rho_{max}$, we can find this positive root efficiently by Newton's method. If we denote $p(t)$ as the quartic function in Eq. (30), it is worthwhile to note that $p(0) = c_0 < 0$ if $\rho < \rho_{max}$. Therefore the initial point α_0 can be chosen such that $p(\alpha_0) > 0$ (such α_0 always exists since $p(\alpha) \to \infty$ when $\alpha \to \infty$). Once the Newton's method fails, e.g., there is a $\bar{\alpha} \in (\alpha, \alpha_0)$ such that $p(\bar{\alpha}) > 0$ and $p'(\bar{\alpha}) = 0$, we switch to the more stable solver via eigendecomposition of a 4×4 matrix (**?**). However, Newton's method never fails in our experiments and a few steps (less than 10) usually result in a very accurate solution ($|p(\alpha_t)| < 1e - 18$ where α_t is the value returned by the Newton's method after t iterations).

8. Proof of Theorem (3)

We are trying to solve:

$$\min_{\mathbf{u}} h(\mathbf{u}) = \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + \rho|u_2 - u_1| \tag{31}$$

where $\rho = \lambda/\gamma$.

Proof. $\mathbf{u}^*$ is the optimal solution of problem (31) if and only if $0 \in \partial_{\mathbf{u}} h(\mathbf{u}^*)$, i.e.,

$$\mathbf{u}^* \in \mathbf{w} + v\rho\mathbf{g} \tag{32}$$

where $\mathbf{g} = (1, -1)^T$ and

$$v \in \begin{cases} 1, & \text{if } u_2 > u_1 \\ -1, & \text{if } u_2 < u_1 \\ [-1, 1], & \text{if } u_2 = u_1 \end{cases} \tag{33}$$

We write (32) in componentwise:

$$u_1^* = w_1 + v\rho \text{ and } u_2^* = w_2 - v\rho.$$

Since $|v| \leq 1$, then if

$$w_1 - \rho > w_2 + \rho, \text{ i.e., } v = -1$$

we can conclude that $u_1^* > u_2^*$. By (33), v has to be -1. Therefore

$$u_1^* = w_1 - v \text{ and } u_2^* = w_2 + v.$$

Similarly, when

$$w_1 + \rho < w_2 - \rho, \text{ i.e., } v = 1$$

we can see that u_1^* must be smaller than u_2^*. Therefore (33), v has to be 1 and thus

$$u_1^* = w_1 + v \text{ and } u_2^* = w_2 - v.$$

Otherwise, we have

$$w_2 - 2\rho \leq w_1 \leq w_2 + 2\rho,$$

which implies

$$-1 \leq \frac{w_2 - w_1}{2\rho} \leq 1.$$

Therefore if we set $v = \frac{w_2 - w_1}{2\rho}$, we have

$$w_1 + v\rho = w_2 - v\rho,$$

i.e.,

$$u_1^* = u_2^* = \frac{w_1 + w_2}{2}.$$

The results above can be summarized as:

$$\mathbf{u}^* = \begin{cases} (w_1 - \rho, w_2 + \rho)^T, & \text{if } w_1 > w_2 + 2\rho \\ (w_1 + \rho, w_2 - \rho)^T, & \text{if } w_1 < w_2 - 2\rho \\ (\frac{w_1+w_2}{2}, \frac{w_1+w_2}{2})^T, & \text{otherwise} \end{cases} \tag{34}$$

which completes the proof. □