Supplementary material

1. PROOF OF LEMMA 1

Lemma 1. For a differential function $f(\boldsymbol{x})$ whose gradient is $\nabla f(\boldsymbol{x})$, we define its epigraph set as $C_f = \{\bar{\boldsymbol{x}} = [\boldsymbol{x}^{\mathrm{T}}\ q]^{\mathrm{T}} \in \mathbb{R}^{N+1} : f(\boldsymbol{x}) \leq q\}$. Given an initial point $\bar{\boldsymbol{u}}_0 = [\boldsymbol{u}^{\mathrm{T}}\ 0]^{\mathrm{T}}$, the jth ($j \geq 1$) projection onto the epigraph set is denoted as $\bar{\boldsymbol{x}}_j = [\boldsymbol{u}_j^{\mathrm{T}}\ f(\boldsymbol{u}_j)]^{\mathrm{T}}$, then,

$$\boldsymbol{u}_j = \boldsymbol{u}_{j-1} - \frac{f(\boldsymbol{u}_{j-1})}{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1} \nabla f(\boldsymbol{u}_{j-1}) \tag{1}$$

where $\boldsymbol{u}_0 = D_\varepsilon[\boldsymbol{x}^{(t-1)}; \sigma]$.

Proof: In the following, given an initial point $\bar{\boldsymbol{u}}_0 = [\boldsymbol{u}^{\mathrm{T}}\ 0]^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{u} \\ 0 \end{bmatrix}$ and a differential function $f(\boldsymbol{x})$ whose gradient is $\nabla f(\boldsymbol{x})$, we derive its projection onto the epigraph set C_f. For the jth projection, we have the initial point $\bar{\boldsymbol{u}}_{j-1} = \begin{bmatrix} \boldsymbol{u}_{j-1} \\ 0 \end{bmatrix}$. From Fig. 1 (see text), one can denote the coordinate of $\bar{\boldsymbol{x}}_{j-1}$ as $\begin{bmatrix} \boldsymbol{u}_{j-1} \\ f(\boldsymbol{u}_{j-1}) \end{bmatrix}$. Therefore, we can derive the equation of the supporting hyperplane at this point as follows

$$d(\boldsymbol{p}, q) = \nabla f(\boldsymbol{u}_{j-1})(\boldsymbol{p} - \boldsymbol{u}_{j-1}) + f(\boldsymbol{u}_{j-1}) - q = 0 \tag{2}$$

where $\boldsymbol{p}$ and q represent the coordinates in horizontal ordinate axis and vertical ordinate axis defined in Fig. 1 respectively. The normal vector $\vec{\boldsymbol{w}}$ of the supporting hyperplane is $(\nabla f(\boldsymbol{u}_{j-1}), -1)$, and its unit version $\vec{\boldsymbol{e}}$ is defined as $\left(\frac{\nabla f(\boldsymbol{u}_{j-1})}{\sqrt{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1}}, -\frac{1}{\sqrt{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1}}\right)$. Hence, the distance between any point $\begin{bmatrix} \boldsymbol{p}_x \\ q_y \end{bmatrix}$ in the space and the supporting hyperplane can be written as

$$r = \frac{d(\boldsymbol{p}_x, q_y)}{||\vec{\boldsymbol{w}}||_2}. \tag{3}$$

Now, we derive the orthogonal projection of $\bar{\boldsymbol{u}}_{j-1} = \begin{bmatrix} \boldsymbol{u}_{j-1} \\ 0 \end{bmatrix}$ onto the supporting hyperplane, i.e., the projection point $\vec{\boldsymbol{w}}_j = \begin{bmatrix} \boldsymbol{u}_j \\ q_j \end{bmatrix}$. Denoting the original point $\bar{\boldsymbol{o}} = \begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix}$ in Fig. 1, we have

$$\overrightarrow{\bar{\boldsymbol{o}}\vec{\boldsymbol{w}}_j} = \overrightarrow{\bar{\boldsymbol{o}}\bar{\boldsymbol{u}}_{j-1}} - \overrightarrow{\vec{\boldsymbol{w}}_j\bar{\boldsymbol{u}}_{j-1}} = \overrightarrow{\bar{\boldsymbol{o}}\bar{\boldsymbol{u}}_{j-1}} - \frac{d(\boldsymbol{u}_{j-1}, 0)}{||\vec{\boldsymbol{w}}||_2}\vec{\boldsymbol{e}}. \tag{4}$$

Since $d(\boldsymbol{u}_{j-1}, 0) = f(\boldsymbol{u}_{j-1})$ and $||\vec{\boldsymbol{w}}||_2 = \sqrt{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1}$, we have

$$\overrightarrow{\bar{\boldsymbol{o}}\vec{\boldsymbol{w}}_j} = \overrightarrow{\bar{\boldsymbol{o}}\bar{\boldsymbol{u}}_{j-1}} - \frac{f(\boldsymbol{u}_{j-1})}{\sqrt{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1}}\vec{\boldsymbol{e}}. \tag{5}$$

We know that $\overrightarrow{\bar{\boldsymbol{o}}\vec{\boldsymbol{w}}_j} = \begin{bmatrix} \boldsymbol{u}_j \\ q_j \end{bmatrix}$ and $\overrightarrow{\bar{\boldsymbol{o}}\bar{\boldsymbol{u}}_{j-1}} = \begin{bmatrix} \boldsymbol{u}_{j-1} \\ 0 \end{bmatrix}$, hence, we get

$$\boldsymbol{u}_j = \boldsymbol{u}_{j-1} - \frac{f(\boldsymbol{u}_{j-1})}{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1} \nabla f(\boldsymbol{u}_{j-1}), \tag{6}$$

$$q_j = \frac{f(\boldsymbol{u}_{j-1})}{||\nabla f(\boldsymbol{u}_{j-1})||_2^2 + 1}. \tag{7}$$

We have obtained the projection point $\vec{\boldsymbol{w}}_j = \begin{bmatrix} \boldsymbol{u}_j \\ q_j \end{bmatrix}$. Therefore, the jth ($j \geq 1$) projection onto the epigraph set is $\bar{\boldsymbol{x}}_j = [\boldsymbol{u}_j^{\mathrm{T}}\ f(\boldsymbol{u}_j)]^{\mathrm{T}}$ where $\boldsymbol{u}_j$ is updated as Eqn. (1).

This completes the proof.

2. PROOF OF THE BOUNDED PROPERTY OF BM3D

Bound of the BM3D denoiser. *We assume that there exists two monotone decreasing functions $f(\sigma)$ and $\varepsilon(\sigma)$ admitting $\lim_{\sigma\to 0} f(\sigma) = p$ and $\lim_{\sigma\to 0} \varepsilon(\sigma) = 0$. For any input $\boldsymbol{x} \in \mathbb{R}^N$ whose elements admit $x_i \in [0, 1]$ and some universal constant $\widetilde{L} = pC_{max}^2$, the BM3D denoiser is bounded such that*

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}} H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \leq [\varepsilon(\sigma)]^2[p - f(\sigma)] \leq \sigma^2\widetilde{L}. \tag{8}$$

Proof: Firstly, we introduce the matrix and frame representations of analysis and synthesis BM3D operations. Let $\boldsymbol{x}_n \in \mathbb{R}^N$ be a noisy image, and $\boldsymbol{R}_i\boldsymbol{x}_n \in \mathbb{R}^n$ be the ith image patch extracted from the noisy image $\boldsymbol{x}_n$. Here, matrix $\boldsymbol{R}_i \in \mathbb{R}^{n\times N}$ represents the operator that extracts the patch from the whole image. The total number of image patches in each group is a fixed number K, and the total number of groups is R. Let $J_r = \{i_{r,1}, \ldots, i_{r,k}\}$ be the set of image patch indexes in the rth group, thus grouping can be defined by $J = \{J_r : r = 1, \ldots, R\}$. Based on these notations, the explicit matrix representation of the BM3D analysis operation can be written as

$$\boldsymbol{w} = \boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{\Psi}_{\mathrm{BM3D}_1} \\ \cdots \\ \boldsymbol{\Psi}_{\mathrm{BM3D}_R} \end{bmatrix} \boldsymbol{x}_n \tag{9}$$

where $\boldsymbol{w}$ represents the joint 3D groupwise spectrum, and $\boldsymbol{\Psi}_{\mathrm{BM3D}_R}$ is defined as

$$\boldsymbol{\Psi}_{\mathrm{BM3D}_R} = \sum_{i\in J_r} \boldsymbol{d}_i \otimes [(\boldsymbol{D}_2 \otimes \boldsymbol{D}_2)\boldsymbol{R}_i] \tag{10}$$

where $\otimes$ represents the Kronecker matrix product. The 3D decorrelating transform is constructed as a separable combination of 2D intrablock and 1D interblock transforms. Here, each block is an image patch. In Eqn. (10), $\boldsymbol{d}_i$ is the ith column of 1D interblock transform $\boldsymbol{D}_1$, and $\boldsymbol{D}_2$ represents the 1D transform that constitutes the separable 2D intrablock transform.

The synthesis matrix is derived similarly. The estimated image is the weighted mean of the groupwize estimation, and the weights are defined as $g_r > 0$. The synthesis representation model for a noisy image can be represented as

$$\boldsymbol{x}_n = \boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{w} = \boldsymbol{W}^{-1}[g_1\boldsymbol{\Phi}_{\mathrm{BM3D}_1}, \cdots, g_R\boldsymbol{\Phi}_{\mathrm{BM3D}_R}]\boldsymbol{w} \tag{11}$$

where $\boldsymbol{W} = \sum_r g_r \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$. The explicit matrix representation of the BM3D synthesis operation can be formulated as

$$\boldsymbol{\Phi}_{\mathrm{BM3D}_R} = \sum_{i\in J_r} \boldsymbol{d}_i^{\mathrm{T}} \otimes [\boldsymbol{R}_i^{\mathrm{T}}(\boldsymbol{D}_2 \otimes \boldsymbol{D}_2)^{\mathrm{T}}]. \tag{12}$$

Remark Based on the definition of the analysis and synthesis BM3D operations, the matrix $\boldsymbol{\Psi}_{\mathrm{BM3D}}$ admits: $\boldsymbol{\Psi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Psi}_{\mathrm{BM3D}} = \sum_r \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i > 0$ and the matrix $\boldsymbol{\Phi}_{\mathrm{BM3D}}$ admits: $\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}} = \boldsymbol{W}^{-2}\sum_r {g_r}^2 \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i > 0$. Therefore, the following equation hold for the underlying or clean image $\boldsymbol{x} \in \mathbb{R}^N$:

$$\boldsymbol{x} = \boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}. \tag{13}$$

Based on the matrix representations of analysis and synthesis BM3D operations, the analysis and synthesis BM3D frames are defined as follows. Definition of analysis BM3D frame: It follows from Remark that rows of $\boldsymbol{\Psi}_{\mathrm{BM3D}}$ constitute frame in $\{\boldsymbol{\Psi}_{\mathrm{BM3D}_k}\}$; Definition of synthesis BM3D frame: It follows from Remark that columns of $\boldsymbol{\Phi}_{\mathrm{BM3D}}$ constitute frame in $\{\boldsymbol{\Phi}_{\mathrm{BM3D}_k}\}$. It follows the Remark that the BM3D frames admit dual property

$$\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}} = \boldsymbol{I}. \tag{14}$$

According to the analysis and synthesis BM3D frames, the BM3D denoiser can be described as

$$D_{\mathrm{BM3D}}(\bullet;\sigma) = \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}(\bullet)] \tag{15}$$

where $H_{\varepsilon(\sigma)}$ is the hard thresholding operator using the thresholding value $\varepsilon(\sigma)$. The hard thresholding operator is defined as $H_{\varepsilon(\sigma)}(u) = u$ if $|u| > \varepsilon(\sigma)$, and $H_{\varepsilon(\sigma)}(u) = 0$, otherwise. The parameter $\varepsilon(\sigma)$ is correlated with the input noise standard deviation σ. In general, the thresholding value $\varepsilon(\sigma)$ is a monotone decreasing function of the input parameter σ, and this function admits $\lim_{\sigma\to 0}\varepsilon(\sigma) = 0$.

Then, in order to analyze the bounded property of the BM3D denoiser, we give the following assumption.

Assumption: The number of coefficients in $H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}^{(t-1)}]$ is denoted as m, and the number of coefficients in $\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}^{(t-1)}$ is p, we assume that there exists a decreasing function $f(\sigma)$ such that $m = f(\sigma)$. The BM3D denoiser is asymptotically invariant in the sense that it ensures $\boldsymbol{x}^{(t-1)} = \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}[\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x}^{(t-1)}]$ as $\sigma \to 0$, i.e., $\lim_{\sigma\to 0} f(\sigma) = p$.

Finally, based on the Eqn. (14) and Eqn. (15), we have

$$\begin{aligned} &||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &= ||\boldsymbol{\Phi}_{\mathrm{BM3D}}\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &\le ||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2||\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2. \end{aligned} \tag{16}$$

Here, $||\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x} - H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2$ represents the square of the l_2 norm of the filtered spectrum or filtered coefficient vector. Among these lost coefficients, the bound of these filtered coefficients is the threshoulding value $\varepsilon(\sigma)$. Therefore,

$$\begin{aligned} &||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \\ &\le ||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2[\varepsilon(\sigma)]^2[p - f(\sigma)] \end{aligned} \tag{17}$$

where $[p - f(\sigma)]$ is the number of the lost or filtered coefficients. Due to $||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 = \lambda_{max}(\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{BM3D}})$ (here, $\lambda_{max}(\bullet)$ represents the maximum eigenvalue), we have

$$||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 = \lambda_{max}\left(\frac{\sum_r {g_r}^2 \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i}{\sum_r g_r \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i \bullet \sum_r g_r \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i}\right). \tag{18}$$

Letting $\boldsymbol{A} = \sum_r {g_r}^2 \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$ and $\boldsymbol{B} = \sum_r g_r \sum_{i\in J_r} \boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_i$, these two matrices are diagonal matrices. In general, it is assumed that for each pixel there is at least one image patch containing the pixel and entering in some group. Moreover, the weights g_r are in the range of $(0, 1]$. Based on these assumptions, the element values in $\boldsymbol{A}$ are in the range of $[{g_r}^2, {g_r}^2 KR]$, and the element values in $\boldsymbol{B}$ are in the range of $[g_r, g_r KR]$. Therefore, the matrix $\frac{\boldsymbol{A}}{\boldsymbol{B}\boldsymbol{B}}$ is a diagonal matrix whose diagonal elements are in the range of $[\frac{1}{KR}, 1]$. The maximum eigenvalue of the $\boldsymbol{\Phi}_{\mathrm{BM3D}}^{\mathrm{T}}\boldsymbol{\Phi}_{\mathrm{BM3D}}$ is 1, i.e.,

$$||\boldsymbol{\Phi}_{\mathrm{BM3D}}||_2^2 \le 1. \tag{19}$$

Substituting Eqn. (19) into Eqn. (17), we have

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [\varepsilon(\sigma)]^2[p - f(\sigma)]. \tag{20}$$

According to $p - f(\sigma) \le p$, the upper bound of Eqn. (20) can be further determined as follows:

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [\varepsilon(\sigma)]^2 p. \tag{21}$$

Due to the $\varepsilon(\sigma)$ is a monotone decreasing function of σ, there must exist a function $C(\sigma)$ satisfying $\varepsilon(\sigma) = C(\sigma)\sigma$. For a practical BM3D denoiser, the maximum value of $C(\sigma)$ must exist, and we term it as C_{max}. Then Eqn. (21) can be recast as

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le [C(\sigma)\sigma]^2 p \le C_{max}^2\sigma^2 p. \tag{22}$$

Let $\widetilde{L} = pC_{max}^2$, we have

$$||\boldsymbol{x} - \boldsymbol{\Phi}_{\mathrm{BM3D}}H_{\varepsilon(\sigma)}(\boldsymbol{\Psi}_{\mathrm{BM3D}}\boldsymbol{x})||_2^2 \le \sigma^2\widetilde{L}. \tag{23}$$

This completes the proof.