IISE Transactions on Healthcare Systems Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uhse21

Image decomposition-based sparse extreme pixel-level feature detection model with application to medical images

Geet Lahoti, Jialei Chen, Xiaowei Yue, Hao Yan, Chitta Ranjan, Zhen Qian, Chuck Zhang & Ben Wang

To cite this article: Geet Lahoti, Jialei Chen, Xiaowei Yue, Hao Yan, Chitta Ranjan, Zhen Qian, Chuck Zhang & Ben Wang (2021): Image decomposition-based sparse extreme pixel-level feature detection model with application to medical images, IISE Transactions on Healthcare Systems Engineering, DOI: 10.1080/24725579.2021.1910599

To link to this article: https://doi.org/10.1080/24725579.2021.1910599

Published online: 28 May 2021.

IISE TRANSACTIONS ON HEALTHCARE SYSTEMS ENGINEERING
https://doi.org/10.1080/24725579.2021.1910599

Image decomposition-based sparse extreme pixel-level feature detection model with application to medical images

Geet Lahoti[a], Jialei Chen[a], Xiaowei Yue[b], Hao Yan[c], Chitta Ranjan[d], Zhen Qian[e], Chuck Zhang[a], and Ben Wang[f]

[a]H. Milton Stewart School of Industrial & Systems Engineering, and Georgia Tech Manufacturing Institute, Georgia Institute of Technology, Atlanta, GA, USA; [b]Grado Department of Industrial & Systems Engineering, Virginia Tech, Virginia, USA; [c]School of Computing, Informatics, & Decision Systems Engineering, Arizona State University, AZ, USA; [d]ProcessMiner, Atlanta, GA, USA; [e]Medical AI Lab, Tencent America, Palo Alto, CA, USA; [f]Georgia Tech Manufacturing Institute, H. Milton Stewart School of Industrial & Systems Engineering, and the School of Materials Science & Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

Pixel-level feature detection from images is an essential but challenging task encountered in domains such as detecting defects in manufacturing systems and detecting tumors in medical imaging. Often, the real image contains multiple feature types. The types with higher pixel intensities are termed as positive (extreme) features and the ones with lower pixel intensities as negative (extreme) features. For example, when planning a medical treatment, it is important to identify, (a) calcification (a pathological feature which can result in a post-surgical complications) as positive features, and (b) soft tissues (organ morphology, knowledge of which can support pre-surgical planning) as negative features, from a preoperative computed tomography image of the human heart. However, this is not an easy task because (a) conventional segmentation techniques require manual intervention and post-processing, and (b) existing automatic approaches do not distinguish positive features from negative. In this work, we propose a novel, automatic image decomposition-based sparse extreme pixel-level feature detection model to decompose an image into mean and extreme features. To estimate model parameters, a high-dimensional least squares regression with regularization and constraints is utilized. An efficient algorithm based on the alternating direction method of multipliers and the proximal gradient method is developed to solve the large-scale optimization problem. The effectiveness of the proposed model is demonstrated using synthetic tests and a real-world case study, where the model exhibits superior performance over existing methods.

KEYWORDS

Image segmentation; feature detection; medical image analysis; high-dimensional regression; large-scale convex optimization

1. Introduction

Image segmentation is an important task in computer vision. It involves partitioning an image into multiple segments. Each segment can be regarded as consisting of a group of pixel-level features (Sanjay-Gopal & Hebert, 1998). Therefore, a segmentation task can also be regarded as pixel-level feature detection, which is encountered in domains such as manufacturing and medicine. For example, (a) in manufacturing systems, an image of a product is sensed for product inspection; (b) in medicine, an image of a body part is acquired for disease diagnosis and treatment planning. Both product inspection and disease diagnosis require detecting relevant pixel-level features from the image. In the manufacturing example, feature detection involves extracting defects or anomalies (Yan et al., 2017), whereas in the case of medical imaging, it involves segmenting pathological structures (Chen et al., 2019). Since images are high-dimensional and exhibit complex spatial structure, feature detection is challenging. Moreover, image sensing and acquisition systems introduce measurement noise that reduces the contrast between the background and features, thus further complicating the feature detection task (Yan et al., 2017).

In medicine, image segmentation or feature detection is of particular importance (Bradshaw et al., 2013; Gaw et al., 2018; Wang, 2015). Consider aortic stenosis (AS), which is one of the most common yet severe valvular heart diseases. Transcatheter aortic valve replacement (TAVR) is a less-invasive treatment option for AS patients who have a high risk of open-heart surgery. TAVR procedure involves implanting a bioprosthetic aortic valve. Major post-procedural complications of TAVR are the paravalvular leakage (PVL), i.e. the blood flow leakage around the implanted artificial valve (Qian et al., 2017; Wang et al., 2018), and the over-stretching in the aortic tissues introduced due to the implant. For patients undergoing TAVR, computed tomography (CT) image (see Figure 1) is usually taken before the surgery, as an important visualization of the *contrast-enhanced blood pool* (i.e. the moderate intensity region in the CT image), the *calcification* (i.e. the high-intensity region), and the *soft tissues* (i.e. the low-intensity region).

CONTACT Geet Lahoti glahoti6@gatech.edu

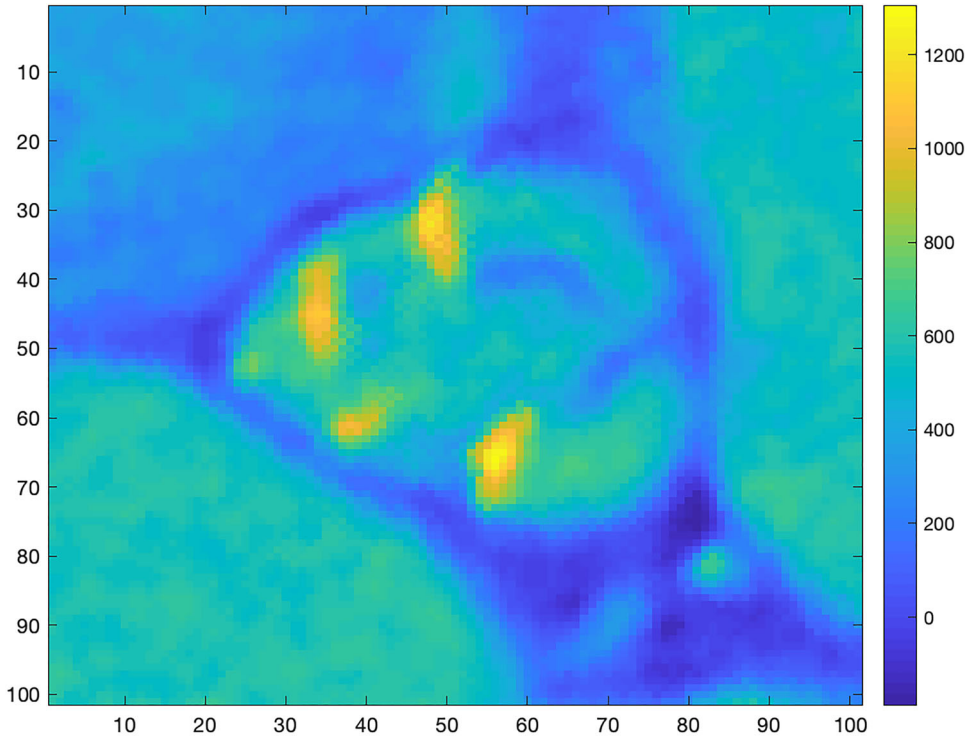

Figure 1. An example of a 2-dimensional computed tomographic image, showing the soft tissues (blue), the calcification (bright yellow), and the blood pool (bluish yellow).

For the clinical decision making and pre-procedural planning, the identification of the soft tissues and the calcification from the CT image is essential. From the identified soft tissues, the structure of the aortic valve can be extracted. The knowledge of the structure of the patient's aortic valve then helps in selecting an appropriately-sized artificial aortic valve. If the artificial valve is not chosen appropriately, post-TAVR complications may follow. For example, if the implanted artificial valve is too small, severe PVL may occur, whereas an oversized artificial valve may introduce high stress in the aortic tissues. Another clinically important factor is the calcification present in the aortic annulus region. If the calcification volume and its distribution are effectively quantified directly from the CT image, physicians can better assess the patient's condition and therefore patient-specific treatment planning can be conducted to mitigate the post-TAVR PVL amount.

Furthermore, due to the advancement in 3 D-printing technology (Chen et al., 2018), treatment planning of TAVR can also be assisted by the 3 D-printed, patient-specific virtual aortic valve (Chen et al., 2018; Qian et al., 2017). One of the crucial steps in this approach is to extract a printable digital model (containing both the soft tissues and the calcification) from CT image (Chen et al., 2020), which again requires precise estimation and detection of the aortic valve region and calcification.

Hence, in order to better plan the TAVR surgery and reduce post-TAVR complications, a precise segmentation and extraction of the soft tissues (aortic valve's structure) and the calcification should be conducted based on the CT image. However, this is a challenging task, mainly due to the complexity involved with the CT data. Currently, the most common practice is to consult radiologists who identify the aortic valve and calcification regions, but this can be both time-consuming and sometimes inaccurate. To reduce human toil, image segmentation techniques such as global thresholding and multi-level thresholding are sometimes used (Tuncay et al., 2015). Since global thresholding simply performs image binarization using a single threshold, experienced radiologists may be required to extract the desired region from the image. As for multi-level thresholding, several levels of thresholding need to be carried out to partition the image to finally obtain desired regions. However, thresholding-based segmentation approaches do not consider spatial correlation among image pixels. As a result, the selected features may not form connected regions and have pathological meaning. Moreover, due to a large variation in the pathological condition, different patients often have completely different contrast agent concentrations and distributions of pixel intensity in the CT image. As a result, every patient requires a dedicated analysis.

In this work, we propose a novel, automatic sparse extreme pixel-level feature detection (PFD) model to identify the soft tissues and the calcification. Our model decomposes an image into four components: mean, positive extreme features, negative extreme features, and noise. The mean part is assumed to be smooth (differentiable), containing only the overall tone of the CT image and can be approximated using a basis. The extreme features which refer to both the calcification (i.e. a high-intensity region, above the mean) and the soft tissues (i.e. a low-intensity region, below the mean), are assumed to be sparse or can be sparsely represented. Their spatial structures can be modeled using different bases. Moreover, appropriate constraints on basis coefficients of

features are imposed to keep positive and negative features separate, thereby overcoming identifiability issues. To estimate the parameters in the PFD model with both sparsity regularization and constraints on bases coefficients, a convex optimization problem is framed, which can be efficiently solved using alternating direction method of multipliers (ADMM) with proximal gradient method (PGM). Using the proposed PFD model, both the calcification and the soft tissues can be identified with high accuracy, thereby permitting the follow-up diagnosis procedure and the treatment planning of TAVR. Moreover, it is worth emphasizing that the PFD model and the associated algorithm can be applicable for other cases as well, where pixel-level feature/anomaly detection from images is required.

The structure of the paper is as follows. The relevant literature is reviewed in section 2. The proposed modeling and estimation framework is illustrated in section 3. The simulation study to evaluate the proposed framework is presented in section 4. In addition, a real data set is used to evaluate the performance of the proposed framework in section 5. Finally, conclusions and scope for future work are discussed in section 6.

2. Literature review

Considerable research has been conducted in the area of medical image segmentation. The most commonly used methods are thresholding, edge detection, and region growing. Thresholding methods are further classified into two categories – global and local (multi-level). In global thresholding, an image is divided into two regions: one having pixel intensities less than a threshold and another having intensities greater than the threshold (Otsu, 1979). The threshold can be manually or automatically selected. Using only a single threshold value, finding precise delineation of region boundaries remains a difficult task. In local (multi-level) thresholding, an image can be segmented into different regions using multiple thresholds (Sahoo et al., 1988). The major drawback of thresholding techniques is that these do not consider the spatial relationships among features present in an image. Due to this, these techniques are very sensitive to noise.

Edge detection methods utilize the gradient information to identify edges in the image. Several edge detection operators such as Sobel, Laplacian, and Canny operators are commonly used (Sharma & Aggarwal, 2010). However, edge detection methods are not robust to noise. Moreover, the these methods often output discontinuous pixels which may not form a continuous curve or a connected region. As a result, most of these methods require a post-processing step.

Region-growing methods start with a single seed point or a set of seed pixels and continue exploring neighboring pixels recursively. Based on certain criteria such as uniformity and connectivity, it is determined whether the neighboring pixels should be added to the region (Adams & Bischof, 1994). However, these methods typically require a set of initial seed pixels, where the (sparse) extreme regions can be grown from.

Active contours, which are also commonly called *snakes*, are used extensively for image segmentation (Leymarie & Levine, 1993). Snake is an energy-minimizing curve that can move through the spatial domain of an image under the guidance of internal forces designed to hold the curve together and keep it from bending too much, and external forces designed to draw curves toward the desired features (Kass et al., 1988; Xu & Prince, 1998). Researchers have developed a semi-automatic segmentation algorithm to extract aortic valve and calcification based on a thresholding technique and snake algorithm (Tuncay et al., 2015). Their algorithm involves multiple steps: (a) the Sinus of Valsalva on the multiplanar reconstruction image is manually detected and cropped off; (b) the cropped region is binarized by global thresholding; (c) the manual intervention is again required to place the initial seed points and create an initial contour on which gradient vector flow snake (Xu & Prince, 1998) algorithm is applied to extract aortic valve area. To detect calcification, a heuristic-based thresholding approach utilizing a histogram of image pixel intensity is used. The major drawback of this approach is that it requires manual intervention and is too case-specific.

Atlas-based segmentation techniques have also been used for medical image segmentation. In this type of technique, an atlas or a lookup table containing information on different features is compiled with the help of experts who label a set of existing images (Sharma & Aggarwal, 2010). Atlas template and target image are registered so that atlas labels can be propagated to the target image. Deformable subdivision surface fitting is another segmentation technique in which a subdivision surface, a smooth boundary surface controlled by a coarse mesh with local support, is fitted to the image data (Kitslaar et al., 2015). Researchers have proposed a multi-step approach combining atlas-based segmentation and deformable subdivision surface fitting to extract aortic root (Gao et al., 2016). However, atlas-based segmentation has high computational cost, while in deformable subdivision surface fitting, the initial model greatly impacts the result.

Statistical shape model (SSM) is another segmentation technique that involves learning the shape of the region of interest using prior knowledge in the form of a training data set (Cootes et al., 1995). The major limitation of this method is that it requires many uniform training samples. Moreover, if the test sample strongly deviates from the training data set, the method is prone to failure. In pathological settings, such a scenario is highly likely.

In optimization-based segmentation techniques, the task of image partitioning is formulated as an energy minimization problem where the objective function mainly consists of two terms: a data term and a regularization term. While the data term reflects the strength of association between a pixel and its label, the regularization term includes priors on the space of feasible solutions and deviations from the priors (Nieuwenhuis et al., 2013; Wang et al., 2013). For more information on optimization-based are referred to (Nosrati and Hamarneh, 2016).

Recently, an image decomposition-based model, known as smooth-sparse decomposition (SSD), has been developed to detect anomalies in noisy images with smooth

backgrounds (Yan et al., 2017). The SSD model decomposes an image into a smooth mean, sparse anomalies, and noise. It is cast as a penalized high-dimensional regression problem, which is solved using large-scale optimization techniques. Despite being effective in separating anomalies from a smooth background, it doesn't explicitly allow separating positive features from negative ones. Post-processing SSD results may help in separating positive features from negative ones, but it is not guaranteed that both types of features can be clearly separated. Since SSD uses a single basis for anomalies, the desired separation of positive and negative features may sometimes not be achieved due to identifiability issues faced by the optimization problem. Although it is possible to extend SSD so that different bases for positive and negative anomalies can be used, setting different bases for positive and negative features without considering appropriate constraints on the coefficients of the bases may result into identifiability issues and if applied to our case, the resulting features may not have pathological meaning.

Therefore, in this work, we propose an approach which ensures both positive and negative extreme features, while approximated by the same basis, are clearly identifiable by imposing constraints on the coefficients of the bases. Furthermore, we show how to incorporate different bases for positive and negative extreme features to detect them precisely.

3. Pixel-level feature detection

In this section, we present the proposed PFD model and the optimization algorithm for parameter estimation. The proposed model is flexible in the sense that it can be used to extract either positive or negative extreme features or both from a high-dimensional image. Although in this work we focus on the CT image, the proposed model is generic enough to be applicable for other image types occurring in different domains such as medical imaging and manufacturing systems. Also, the proposed framework can be extended to higher dimensions (e.g. 3-D images).

3.1. Proposed model

Let us consider a 2-D image denoted by $Y \in \mathbb{R}^{n_1 \times n_2}$, where n_1 and n_2 represent dimensions of the image. Since we are interested in detecting the positive and negative features, the following PFD model is proposed to decompose an image into four components:

$$Y = M + A_p + A_n + E, \tag{1}$$

where M is the mean, A_p are the positive features, A_n are the negative features, and E is the error (there is no distributional assumption required). The mean and features can be further expanded using different bases such as B-splines and wavelets. Therefore, the model can be re-written as:

$$\tilde{y} = B\theta + B_p\theta_p + B_n\theta_n + E, \tag{2}$$

where $\tilde{y} = vec(Y)$ and $vec(\cdot)$ is an operator that converts a matrix to a column vector; B, B_p, and B_n are the bases for the mean, the positive extreme features, and the negative extreme features, respectively; and θ, θ_p, and θ_n are the corresponding coefficients. Furthermore, we impose constraints in the form of: $\theta_p > 0$ and $\theta_n < 0$. The model presented in (2) is more flexible than the SSD model that decomposes an image in the following fashion: $Y = B\theta + B_a\theta_a + E$, where B_a is the basis for anomalies and θ_a are the corresponding coefficients with no explicit constraints. In addition, if the SSD model is naively extended to allow different bases to be utilized for modeling the positive and negative features, the extended SSD model would look like: $Y = B\theta + B_p\theta_p + B_n\theta_n + E$. There is a possibility that the different bases may interfere with each other and end up modeling each other's features. Therefore, the extended SSD model may not be able to identify both the positive and negative features clearly, resulting in identifiability issues. The proposed PFD model ensures identifiability of the positive and negative features by imposing the necessary constraints on the coefficients.

3.2. Optimization algorithm for parameter estimation

To estimate the model parameters, θ, θ_p, and θ_n, a least squares regression is used. The least squares regression is augmented with L_1 and L_2 penalties to ensure the sparsity of the detected features and smoothness of the estimated mean, respectively. It is also augmented with constraints on the coefficients of the bases. Consequently, the PFD model parameters can be estimated by solving the following optimization problem:

$$\begin{aligned} \underset{\theta, \theta_p, \theta_n}{\text{argmin}} \quad & \|\tilde{e}\|^2 + \lambda\theta^T R\theta + \gamma_p\|\theta_p\|_1 + \gamma_n\|\theta_n\|_1 \\ \text{subject to} \quad & \tilde{y} = B\theta + B_p\theta_p + B_n\theta_n + \tilde{e} \\ & \theta_p > 0 \\ & \theta_n < 0, \end{aligned} \tag{3}$$

where $\|\cdot\|_1$ and $\|\cdot\|$ are L_1 and L_2 norm operators, respectively; λ, γ_p, and γ_n are parameters to be tuned by the user; R is the roughness matrix; and $\tilde{y} = vec(Y), B = B_2 \otimes B_1, B_p = B_{p,2} \otimes B_{p,1}, B_n = B_{n,2} \otimes B_{n,1}, \tilde{e} = vec(E)$, and $\otimes$ is the tensor product. Here, B_i, $B_{p,i}$ and $B_{n,i}$ are the bases along i-th direction ($i=1$ and $i=2$ correspond to x and y directions, respectively). It should be noted that the size of the B is determined by its constituting bases – B_2 and B_1. Suppose the size of the B_i is $n_i \times k_{\mu_i}$, in which k_{μ_i} is the number of basis in the i-th direction, then the size of the B is given by $n_1 n_2 \times k_{\mu_1} k_{\mu_2}$. Similarly, suppose that the $B_{p,i}$ is of size $n_i \times k_{p_i}$, then the size of the B_p is $n_1 n_2 \times k_{p_1} k_{p_2}$.

3.2.1. Optimization Algorithm for PFD

To solve the PFD problem in (3), we can use general convex optimization solvers such as the interior point method (Nesterov & Nemirovskii, 1994) since the loss function is convex. However, the interior point method is often slow for high-dimensional data problems and hence cannot be used for cases such as medical images which can be large due to higher resolutions that are today achieved by

advanced medical imaging technology. In this section, we propose an efficient algorithm to solve the PFD problem.

First, we show that the PFD problem in (3) can be reduced to a constrained weighted least absolute shrinkage and selection operator (LASSO) (Gaines et al., 2018; Zou, 2006), as shown in Appendix A. The PFD problem is equivalent to a constrained weighted LASSO problem in the form of:

$$\begin{aligned} \underset{\theta_p,\,\theta_n}{\text{argmin}} \quad & (\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ & +\gamma_p\|\theta_p\|_1+\gamma_n\|\theta_n\|_1 \\ \text{subject to} \quad & \theta_p>0 \\ & \theta_n<0, \end{aligned} \tag{4}$$

where $H=B(B^TB+\lambda R)^{-1}B^T$.

The constrained weighted LASSO problem in (4) can be solved using quadratic programming (Gaines et al., 2018). However, that may not be efficient in the high-dimensional data setting (Yan et al., 2017). Therefore, an efficient algorithm needs to be developed to solve this problem. We utilize the alternating direction method of multipliers (ADMM) algorithm, which has garnered renewed popularity in statistics and machine learning applications recently (Boyd et al., 2010). This algorithm is easy to implement and suitable for distributed computing setting as well.

To solve the constrained weighted LASSO problem in (4) using the ADMM, it is re-written as follows:

$$\begin{aligned} \underset{\theta_p,\,\theta_n}{\text{argmin}} \quad & f(\theta_p,\theta_n)+g(z_p,z_n) \\ \text{subject to} \quad & \theta_p-z_p=0 \\ & \theta_n-z_n=0, \end{aligned} \tag{5}$$

where $f(\theta_p,\theta_n)=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n)+\gamma_p\|\theta_p\|_1+\gamma_n\|\theta_n\|_1$ and $g(z_p,z_n)=\begin{cases}0 & \text{if } (z_p,z_n)\in\mathcal{C}\\ \infty & \text{if } (z_p,z_n)\notin\mathcal{C}\end{cases}$, with $\mathcal{C}=\{(z_p,z_n): z_p>0, z_n<0\}$. The functions $f(\theta_p,\theta_n)$ and $g(z_p,z_n)$ are found to be closed, proper, and convex, as required by the ADMM. Next, the augmented Lagrangian function is formed as shown below:

$$\begin{aligned} \mathcal{L}_\rho(\theta_p,\theta_n,z_p,z_n,y_p,y_n) =& f(\theta_p,\theta_n)+g(z_p,z_n) \\ & +y_p^T(\theta_p-z_p)+y_n^T(\theta_n-z_n) \\ & +\frac{\rho}{2}(\|\theta_p-z_p\|^2+\|\theta_n-z_n\|^2), \end{aligned} \tag{6}$$

where $\rho>0$. The ADMM algorithm employs block coordinate descent to the augmented Lagrangian followed by an update of the dual variables as shown below:

$$\begin{aligned} \left(\theta_p^{(k)},\theta_n^{(k)}\right) &\leftarrow \underset{\theta_p,\theta_n}{\text{argmin}}\ \mathcal{L}_\rho\left(\theta_p,\theta_n,z_p^{(k-1)},z_n^{(k-1)},y_p^{(k-1)},y_n^{(k-1)}\right) \\ \left(z_p^{(k)},z_n^{(k)}\right) &\leftarrow \underset{z_p,z_n}{\text{argmin}}\ \mathcal{L}_\rho\left(\theta_p^{(k)},\theta_n^{(k)},z_p,z_n,y_p^{(k-1)},y_n^{(k-1)}\right) \\ y_p^{(k)} &\leftarrow y_p^{(k-1)}+\rho\left(\theta_p^{(k)}-z_p^{(k)}\right) \\ y_n^{(k)} &\leftarrow y_n^{(k-1)}+\rho\left(\theta_n^{(k)}-z_n^{(k)}\right), \end{aligned} \tag{7}$$

where super-indices (k) and $(k-1)$ denote iteration numbers. Using (6) and letting $u_p=y_p/\rho$ and $u_n=y_n/\rho$, the updates in (7) can be re-written as follows:

$$\begin{aligned} \left(\theta_p^{(k)},\theta_n^{(k)}\right) &\leftarrow \underset{\theta_p,\theta_n}{\text{argmin}}\ f(\theta_p,\theta_n)+\frac{\rho}{2}\Big(\|\theta_p-z_p^{(k-1)}+u_p^{(k-1)}\|^2 \\ &\qquad +\|\theta_n-z_n^{(k-1)}+u_n^{(k-1)}\|^2\Big) \\ \left(z_p^{(k)},z_n^{(k)}\right) &\leftarrow \underset{z_p,z_n}{\text{argmin}}\ g(z_p,z_n)+\frac{\rho}{2}\Big(\|\theta_p^{(k)}-z_p+u_p^{(k-1)}\|^2 \\ &\qquad +\|\theta_n^{(k)}-z_n+u_n^{(k-1)}\|^2\Big) \\ u_p^{(k)} &\leftarrow u_p^{(k-1)}+\theta_p^{(k)}-z_p^{(k)} \\ u_n^{(k)} &\leftarrow u_n^{(k-1)}+\theta_n^{(k)}-z_n^{(k)}. \end{aligned} \tag{8}$$

Next, we find the solutions for $(\theta_p^{(k)},\theta_n^{(k)})$ and $(z_p^{(k)},z_n^{(k)})$ update problems. First, consider the $(\theta_p^{(k)},\theta_n^{(k)})$ update problem. Substituting for $f(\theta_p,\theta_n)$ and re-arranging, the problem reduces to the following weighted LASSO problem:

$$\begin{aligned} \underset{\theta_p,\,\theta_n}{\text{argmin}} \quad & (\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ & +\frac{\rho}{2}(\|\theta_p-z_p^{(k-1)}+u_p^{(k-1)}\|^2 \\ & +\|\theta_n-z_n^{(k-1)}+u_n^{(k-1)}\|^2)+\gamma_p\|\theta_p\|_1+\gamma_n\|\theta_n\|_1. \end{aligned} \tag{9}$$

The weighted LASSO problem in (9) is first solved for θ_p while keeping $\theta_n=\theta_n^{(k-1)}$. Therefore, the weighted LASSO problem in (9) simplifies into following:

$$\begin{aligned} \underset{\theta_p}{\text{argmin}} \quad & (\tilde{y}-B_p\theta_p-B_n\theta_n^{(k-1)})^T(I-H)(\tilde{y}-B_p\theta_p \\ & -B_n\theta_n^{(k-1)})+\frac{\rho}{2}\|\theta_p-z_p^{(k-1)}+u_p^{(k-1)}\|^2+\gamma_p\|\theta_p\|_1. \end{aligned} \tag{10}$$

The LASSO in 10 can be solved using least angle regression (LARS) (Efron et al., 2004) and quadratic programming (Tibshirani, 1996). However, these methods are inefficient in the high-dimensional setting (Yan et al., 2017; Yue et al., 2017). Therefore, an efficient algorithm based on the proximal gradient (PG) method (Parikh & Boyd, 2013; Tseng, 2008) is developed.

The PG method is a popular optimization algorithm which can be used to solve a class of optimization problems that involves a sum of a group of convex, possibly non-differentiable functions, in the objective function. In our case, the problem in (10) has an objective function which consists of $F(\theta_p)=(\tilde{y}-B_p\theta_p-B_n\theta_n^{(k-1)})^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n^{(k-1)})+\frac{\rho}{2}\|\theta_p-z_p^{(k-1)}+u_p^{(k-1)}\|^2$ and $G(\theta_p)=\gamma_p\|\theta_p\|_1$. It can be easily seen that $F(\theta_p)$ is convex-differentiable. Since $G(\theta_p)$ is an L_1 norm, it is convex but non-differentiable. The PG method also assumes that the convex differentiable function $F(\theta_p)$ has an L-Lipschitz continuous gradient. We prove in Proposition 1 that $F(\theta_p)$ is indeed

Lipschitz continuous, which then ensures the convergence of PG method.

Proposition 1. $F(\cdot)$ is Lipschitz continuous, i.e. there is a constant $L = 2\|B_p\|_2^2 + \rho$, where $\|\cdot\|_2^2$ represents square of matrix spectral norm, such that the gradient $\nabla F(\cdot)$ satisfies $\|\nabla F(\alpha) - \nabla F(\beta)\| \leq L\|\alpha - \beta\|, \forall \alpha, \beta \in \mathbb{R}$.

Proof of the Proposition 1 can be found in Appendix B.

Consequently, the PG method optimizes the weighted LASSO problem in (10) via an iterative algorithm given by:

$$\begin{aligned}\underset{\theta_p}{\text{argmin}} \quad & F\left(\theta_p^{(k-1)}\right) + \left\langle \theta_p - \theta_p^{(k-1)}, \nabla F\left(\theta_p^{(k-1)}\right)\right\rangle \\ & + \frac{L}{2}\|\theta_p - \theta_p^{(k-1)}\|^2 + \gamma_p\|\theta_p\|_1,\end{aligned} \tag{11}$$

where super-indices (k) and $(k-1)$ denote iteration numbers and $\langle\cdot,\cdot\rangle$ refers to the inner product operator. We prove in Proposition 2 that the PG algorithm has a closed-form solution in each iteration k.

Proposition 2. *The proximal gradient algorithm for the problem in (10), given by* $\theta_p^{(k)} = \text{argmin}_{\theta_p}\left\{F(\theta_p^{(k-1)}) + \langle\theta_p - \theta_p^{(k-1)}, \nabla F(\theta_p^{(k-1)})\rangle + \frac{L}{2}\|\theta_p - \theta_p^{(k-1)}\|^2 + \gamma_p\|\theta_p\|_1\right\}$, *has a closed-form solution in each iteration k, in the form of a soft-thresholding function as given below:*

$$\begin{aligned}\theta_p^{(k)} = S_{\frac{\gamma_p}{L}}\Big(\theta_p^{(k-1)} + \frac{2}{L}B_p^T\Big(\tilde{y} - B\theta^{(t-1)} - B_p\theta_p^{(k-1)} \\ - B_n\theta_n^{(k-1)}\Big) - \rho(\theta_p^{(k-1)} - z_p^{(k-1)} + u_p^{(k-1)})\Big).\end{aligned} \tag{12}$$

Proof of the Proposition 2 is available in Appendix C.

Similarly, θ_n can be updated using the following:

$$\begin{aligned}\theta_n^{(k)} = S_{\frac{\gamma_n}{L}}\Big(\theta_n^{(k-1)} + \frac{2}{L}B_n^T\Big(\tilde{y} - B\theta^{(t-1)} - B_p\theta_p^{(k-1)} \\ - B_n\theta_n^{(k-1)}\Big) - \rho\Big(\theta_n^{(k-1)} - z_n^{(k-1)} + u_n^{(k-1)}\Big)\Big).\end{aligned} \tag{13}$$

Now, a solution to $(z_p^{(k)}, z_n^{(k)})$ update problem is discussed. The following problem needs to be solved:

$$\begin{aligned}\underset{z_p, z_n}{\text{argmin}} \quad & g(z_p, z_n) + \frac{\rho}{2}\Big(\|\theta_p^{(k)} - z_p + u_p^{(k-1)}\|^2 \\ & + \|\theta_n^{(k)} - z_n + u_n^{(k-1)}\|^2\Big).\end{aligned} \tag{14}$$

It is easy to see that the solution to the problem in (14) is given by:

$$(z_p^{(k)}, z_n^{(k)}) \quad \leftarrow \text{proj}_{\mathcal{C}}(\theta_p^{(k)} + u_p^{(k-1)}, \theta_n^{(k)} + u_n^{(k-1)}) \tag{15}$$

To compute the matrix, $H = B(B^TB + \lambda R)^{-1}B^T$, a matrix inversion operation is required, which can be computationally expensive as the size of the image increases. Following (Xiao et al., 2013; Yan et al., 2017), the matrix R is defined as $R = B_2^TB_2 \otimes D_1^TD_1 + D_2^TD_2 \otimes B_1^TB_1 + \lambda D_2^TD_2 \otimes D_1^TD_1$, where D_i is the first order difference matrix in the i – *th* direction. This results in a decomposable projection matrix, $H = H_2 \otimes H_1$, where $H_i = B_i(B_i^TB_i + \lambda R)^{-1}B_i^T$. With this, the algorithm becomes efficient for images because the matrix inversion operator is only applied to lower-dimensional matrices. Hence, the computational complexity of the matrix inversion operation is reduced from $O((k_{\mu_1}k_{\mu_2})^3)$ to $O(k_{\mu_1}^3 + k_{\mu_2}^3)$.

3.2.2. Tuning Parameters Selection – λ, γp, γn, ρ:

Choosing the right values of tuning parameters is important in optimization. There are three tuning parameters, λ, γ_p, and γ_n, associated with the PFD model. The other tuning parameter ρ is associated with the ADMM algorithm.

The optimal λ is selected based on the generalized cross-validation (GCV) criterion (Eilers & Marx, 1996), as given below:

$$\lambda_{optimal} = \underset{\lambda}{\text{argmin}} \ \text{GCV}(\lambda) \tag{16}$$

where $\text{GCV}(\lambda) = \frac{\|Y - H_1(\lambda)(Y - A_p - A_n)H_2(\lambda) - A_p - A_n\|^2/n}{(1 - n^{-1}tr(H(\lambda)))^2}$, $A_p = B_{p,1}X_pB_{p,2}^T$, and $A_n = B_{n,1}X_nB_{n,2}^T$. The $H_i(\lambda)$ involves matrix inversion and it needs to be calculated for multiple values of λ. This can be computationally expensive. Following (Ruppert, 2002; Yan et al., 2017), a series of transformations and operations is used to speed up the computations. First, Cholesky decomposition of $B_i^TB_i$ is calculated giving square matrix Z_i. Next, the eigenvalues and eigenvectors of matrix $Z_i^{-1}D_i^TD_iZ_i^{-1}$ are calculated as follows: $Z_i^{-1}D_i^TD_iZ_i^{-1} = U_i\text{diag}(s_i)U_i^T$. If we compute $V_i = B_iZ_i^{-1}U_i$ prior to the optimization, then in each iteration, $H_i(\lambda)$ can be calculated using $H_i(\lambda) = V_i(I + \lambda\text{diag}(s_i))^{-1}V_i^T$ and $tr(H_i(\lambda)) = \sum_{j=1}^n \frac{1}{(1+\lambda s_{ij})}$ (Yan et al., 2017). It can be easily spotted that computing $H_i(\lambda)$ doesn't involve matrix inversion anymore and this device makes the calculation much more efficient.

Instead of choosing optimal γ_p, γ_n, and ρ using a cross-validation method, we adopt an iterative procedure which dynamically tunes these parameters in each iteration of the ADMM algorithm. The optimal γ_p and γ_n can also be selected using the GCV criterion. However, the GCV leads to a higher false positive rate by selecting more pixels because it can be seen from (16) that the GCV minimizes a residual sum of squares (RSS). However, in anomaly detection/feature detection, the objective is to extract the features and not to achieve a smaller RSS (Yan et al., 2017). Hence, Otsu's method (Otsu, 1979) is utilized to select γ_p^k and γ_n^k individually, as suggested in (Yan et al., 2017). In our algorithm, we utilize the notation $otsu(x)$ to represent the function which outputs $\gamma^{(k)}$ by applying Otsu's method on x.

As far as ρ is concerned, it could be set at a fixed value pre-determined by simulation experiments or based on experience. Another option is to use different parameters $\rho^{(k)}$ for each iteration. Let primal residual be defined as $r_k = \sqrt{\|\Theta_p^{(k)} - Z_p^{(k)}\|^2 + \|\Theta_n^{(k)} - Z_n^{(k)}\|^2}$ and dual residual as $s_k = \sqrt{\|Z_p^{(k)} - Z_p^{(k-1)}\|^2 + \|Z_n^{(k)} - Z_n^{(k-1)}\|^2}$. The parameter update scheme used is as follows:

$$\rho^{(k)} = \begin{cases} \kappa^{incr}\rho^{(k-1)} & \text{if } r^k > \mu s^k \\ \rho^{(k-1)}/\kappa^{decr} & \text{if } r^k < \mu^{-1} s^k \\ \rho^{(k)} & \text{otherwise,} \end{cases} \quad (17)$$

where $\mu > 1, \kappa^{incr} > 1$, and $\kappa^{decr} > 1$ are parameters. We use $\mu = 10$ and $\kappa^{incr} = \kappa^{decr} = 1.025$. This scheme helps in keeping primal and dual residuals close to one another as they both converge to zero. Note that the ADMM convergence theory is applicable only when ρ is kept constant. Therefore, when using the aforementioned scheme, ρ should be fixed after a certain number of iterations.

To solve the PFD problem, the detailed optimization algorithm is given in Algorithm 1. Since the optimization problem is convex, from the property of the ADMM and the PG method, it can easily be shown that the proposed algorithm converges to the global optimum (Boyd et al., 2010; Tseng, 2008). To tune the parameter λ, this algorithm is run for different values of λ and the GCV criterion is calculated for each different value. Whichever value of λ maximizes the GCV criterion is selected as $\lambda_{optimal}$.

Algorithm 1: Optimization algorithm for solving PFD

initialization:

B_1, B_2 # mean bases

$B_{p,1}, B_{p,2}$ # positive features bases

$B_{n,1}, B_{n,2}$ # negative features bases

$Z_i = (B_i^T B_i)^{1/2}$, $i = 1, 2$ # Cholesky decomposition

$U_i \text{diag}(s_i) U_i^T = Z_i^{-1} D_i^T D_i Z_i^{-1}$, $i = 1, 2$ # eigenvalues & eigenvectors

$V_i = B_i Z_i^{-1} U_i$, $i = 1, 2$

$H_i(\lambda) = V_i (I + \lambda diag(s_i))^{-1} V_i^T$, $i = 1, 2$

$L_p = 2\|B_p\|_2^2 + \rho, L_n = 2\|B_n\|_2^2 + \rho$ # Lipschitz constant

$X_p^{(0)} = 0, X_n^{(0)} = 0$, $t = 1, \epsilon = 10^{-6}$

while $\|X_p^{(t)} - X_p^{(t-1)}\| + \|X_n^{(t)} - X_n^{(t-1)}\| > \epsilon$ **do**

update positive and negative features

$A_p^{(t-1)} = B_{p,1} X_p^{(t-1)} B_{p,2}^T$

$A_n^{(t-1)} = B_{n,1} X_n^{(t-1)} B_{n,2}^T$

update mean

$M^{(t)} = H_1 (Y - A_p^{(t-1)} - A_n^{(t-1)}) H_2$

initialization:

$\Theta_p^{(0)} = X_p^{(t-1)}$, $\Theta_n^{(0)} = X_n^{(t-1)}$, $Z_p^{(0)} = 0$, $Z_n^{(0)} = 0$, $U_p^{(0)} = 0$, $U_n^{(0)} = 0$, $k = 1$

while $\|\Theta_p^{(k-1)} - Z_p^{(k-1)}\| + \|\Theta_n^{(k-1)} - Z_n^{(k-1)}\| > \epsilon$ OR $\|Z_p^{(k)} - Z_p^{(k-1)}\| + \|Z_n^{(k)} - Z_n^{(k-1)}\| > \epsilon$ **do**

update positive and negative features

$A_p^{(k-1)} = B_{p,1} \Theta_p^{(k-1)} B_{p,2}^T$

$A_n^{(k-1)} = B_{n,1} \Theta_n^{(k-1)} B_{n,2}^T$

update coefficients

$$\Theta_{pe}^{(k)} = \Theta_p^{(k-1)} + \frac{2}{L_p} B_{p,1}^T (Y - M^{(k)} - A_p^{(k-1)} - A_n^{(k-1)}) B_{p,2} - \rho(\Theta_p^{(k-1)} - Z_p^{(k)} + U_p^{(k)})$$

$$\Theta_{ne}^{(k)} = \Theta_n^{(k-1)} + \frac{2}{L_n} B_{n,1}^T (Y - M^{(k)} - A_p^{(k-1)} - A_n^{(k-1)}) B_{n,2} - \rho(\Theta_n^{(k-1)} - Z_n^{(k)} + U_n^{(k)})$$

$\gamma_p / L_p = Otsu(\Theta_{pe}^{(k)}); \gamma_n / L_n = Otsu(\Theta_{ne}^{(k)})$

$\Theta_p^{(k)} = S_{\gamma_p / L_p}(\Theta_{pe}^{(k)}); \Theta_n^{(k)} = S_{\gamma_n / L_n}(\Theta_{ne}^{(k)})$

$Z_p^{(k)} = \text{proj}_{\mathcal{C}}(\Theta_p^{(k)}); Z_n^k = \text{proj}_{\mathcal{C}}(\Theta_n^{(k)})$

update dual variables

$U_p^{(k)} = U_p^{(k-1)} + \Theta_p^{(k)} - Z_p^{(k)}$

$U_n^{(k)} = U_n^{(k-1)} + \Theta_n^{(k)} - Z_n^{(k)}$

end

$X_p^{(t)} = \Theta_p^{(k)}; X_n^{(t)} = \Theta_n^{(k)}$

End

3.2.3. Basis Selection

Selecting the appropriate bases for the mean and the features is also important in the implementation of the PFD model.

For the smooth mean, a spline or kernel basis can be considered. Hyper-parameters associated with the splines (e.g. number of knots) and kernels (e.g. scale parameter) can be tuned using a criterion such as GCV (Yan et al., 2017).

As far as features are concerned, having some prior information on the shape of features is preferable. Some ideas for selecting basis are as follows: (a) if the features are in the form of small regions scattered over the mean or are thin lines, then one can use identity basis, $B_a = I$; (b) if the features are regions with sharp corners, then linear B-splines are recommended; (c) if the features are regions with curved boundaries, higher order splines such as quadratic and cubic, can be utilized (Yan et al., 2018; Yan et al., 2017). Also, knowing an estimate of the size of the feature can help in selecting the optimal number of knots for a spline basis.

4. Simulation study

In this section, we demonstrate the effectiveness of the proposed PFD model and estimation procedure for feature detection in image data. The main objective here is to show, (a) how the proposed model overcomes identifiability issues often faced by the SSD and (b) the flexibility of the proposed model in terms of allowing different bases to be used for positive and negative extreme regions. The performance of the proposed PFD model and its comparison with the SSD is evaluated using data simulated under different conditions. Two other benchmarks, Nick local thresholding (Khurshid et al., 2009) and global thresholding (Otsu, 1979), are also used for

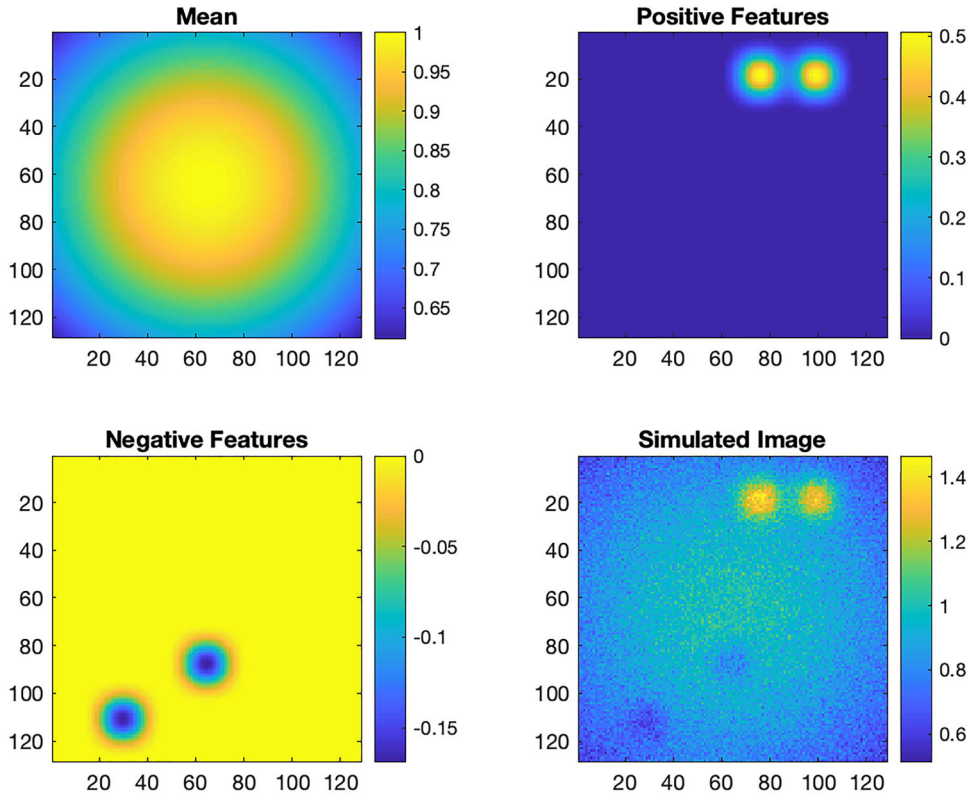

Figure 2. Scenario A – a sample simulated image with $\delta_p = 0.9$ and $\delta_n = -0.3$.

comparison. Two types of scenarios are considered – (a) both positive and negative features are of the same type, (b) positive and negative features are of different types.

Images are simulated using the following model $Y = M + A_p + A_n + E$, where Y is the simulated image, M denotes true mean, A_p and A_n are positive and negative features, respectively, and E is the error. Size of each simulated image is 128×128. In the present case, mean is simulated using $M(x,y) = \exp\left(-\frac{x^2+y^2}{4}\right)$, where $x = \frac{i}{128}, y = \frac{j}{128}, i = j = 1, \ldots, 128$. To simulate random noise, E is sampled from i.i.d $N(0, \phi^2)$ with $\phi = 0.05$.

For both methods – SSD and PFD, the cubic B-spline with 3×3 knots is used for the mean estimation. Since the focus here is to show the applicability of the PFD method when the SSD method faces identifiability issues and is not capable of allowing different bases for features, known bases are used to simulate the features.

The following performance metrics are used to compare the performance of the methods – mean/background recovery root mean square error (e_M) defined as the square root of mean square error of the mean estimator $\hat{M}$: $e_M = \sqrt{\|M - \hat{M}\|^2}$; positive features recovery root mean square error (e_{A_p}) of the positive features estimator $\widehat{A_p}$: $e_{A_p} = \sqrt{\|A_p - \widehat{A_p}\|^2}$; negative features recovery root mean square error (e_{A_n}) of the negative features estimator $\widehat{A_n}$: $e_{A_n} = \sqrt{\|A_n - \widehat{A_n}\|^2}$; features recovery root mean square error: $e_A = \sqrt{\|A_n - \widehat{A_n}\|^2 + \|A_p - \widehat{A_p}\|^2}$; precision defined as the percentage of recovered features by the algorithm that are indeed features; and recall defined as the percentage of the true features detected by the algorithm.

4.1. Scenario A - same type of positive and negative features

In this scenario, both positive and negative features are simulated using the same type of basis B_a, which is chosen as cubic B-spline with 12×12 knots. Specifically, positive features are obtained using $A_p = B_a A_{ps} B_a^T$, where A_{ps} is a sparse matrix with size 13×13 in which two randomly selected entries are replaced with a small number $\delta_p > 0$. Negative features are obtained using $A_n = B_a A_{ns} B_a^T$, where A_{ns} is a sparse matrix of size 13×13 in which two-randomly selected entries are replaced with a small number $\delta_n < 0$. A sample simulated image along with mean, positive features, and negative features is shown in Figure 2.

In the experiments discussed in this scenario, we fix and use $\delta_n = -0.3$. We then try different values for δ_p. The recovery square root mean square error performance metrics for SSD with post-processing (PP) (to separate positive and negative features using a pre-defined threshold as 0), and PFD are compared in Figures 3 and 4. As can be seen from Figure 3, e_M for PFD remains almost constant with an increase in δ_p, whereas that for SSD (PP) becomes higher starting with $\delta_p = 0.9$. The performance metric, e_{A_p}, remains almost constant and same for both the methods as δ_p increases. The metric, e_{A_n}, for PFD remains very low with an increase in δ_p. On the other hand, there is a sudden jump in e_{A_n} for SSD (PP) at and after $\delta_p = 0.9$. This

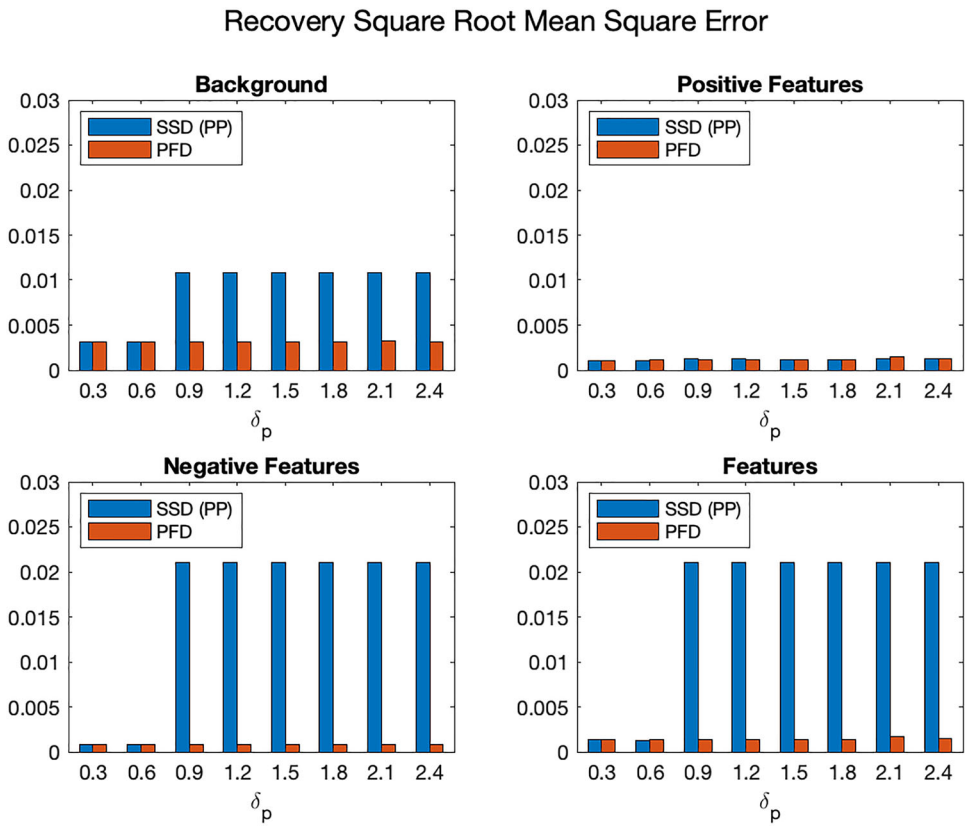

Figure 3. Scenario A – recovery square root mean square error results showing comparison between SSD (PP) and PFD.

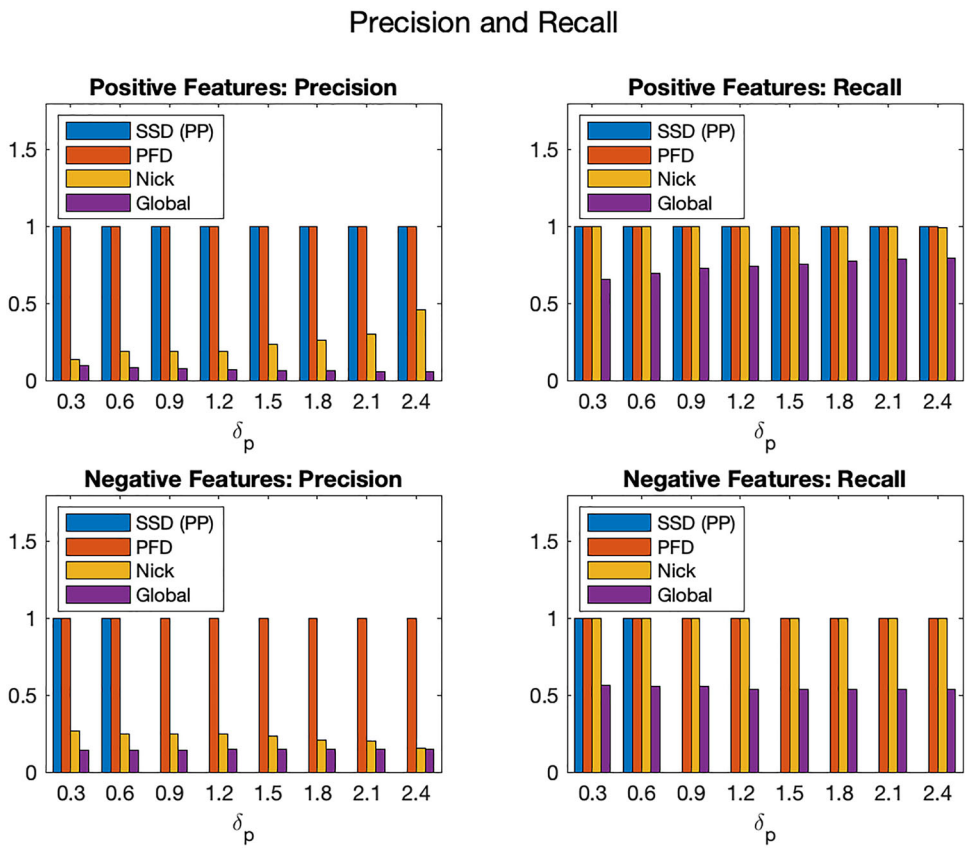

Figure 4. Scenario A – precision and recall results showing comparison among SSD (PP), PFD, Nick local thresholding, and global thresholding.

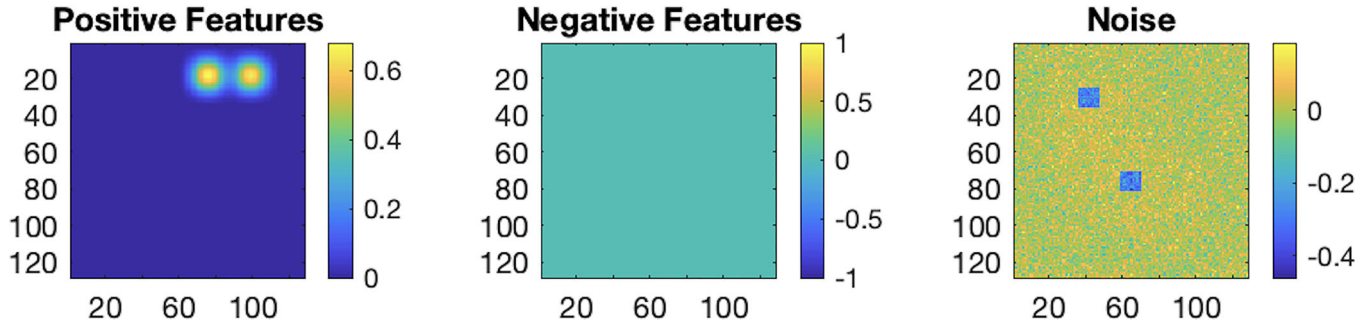

Figure 5. Scenario A – identifiability issue faced by SSD (PP), when $\delta_p = 0.9$.

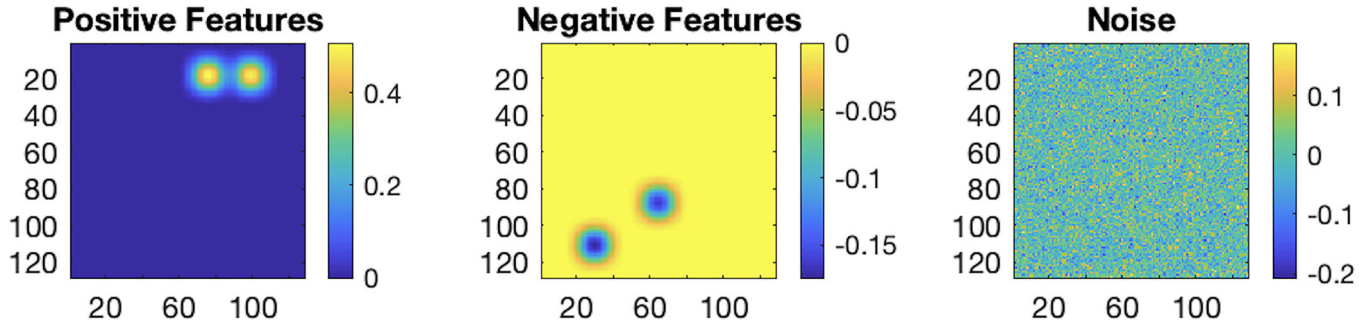

Figure 6. Scenario A – identifiability successfully dealt with PFD, when $\delta_p = 0.9$.

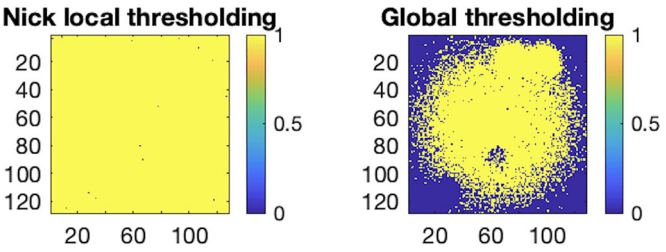

Figure 7. Scenario A – features detected by Nick local thresholding and global thresholding, when $\delta_p = 0.9$.

indicates the onset of the identifiability issue. When feature recovery performance is seen through a combined metric, e_A, PFD clearly outperforms SSD (PP). Next, we discuss the performance in terms of precision and recall as illustrated in Figure 4. PFD maintains perfect precision and recall scores when δ_p varies, whereas SSD (PP) is unable to detect negative features starting with $\delta_p = 0.9$ because it doesn't allow selecting separate thresholds for positive and negative features. This phenomenon is illustrated by Figures 5 and 6. Although Nick local thresholding maintains a perfect recall, its precision is very low in comparison to PFD. Global thresholding shows poor performance with very low precision and recall among all the methods. The sample results for the Nick local and the global thresholding are shown in Figure 7. In this scenario, the PFD clearly outperforms the existing methods in terms of all the performance metrics.

4.2. Scenario B - different type of positive and negative features

In this scenario, positive features are generated using $A_p = B_p A_{ps} B_p^T$, where B_p is a cubic B-spline basis with 12 knots and A_{ps} is a sparse matrix of size 13×13 in which two randomly selected entries are replaced with a small number $\delta_p > 0$. The negative features are generated using $A_n = B_n A_{ns} B_n^T$, where B_n is a linear B-spline basis with 12 knots and A_{ns} is a sparse matrix of size 11×11 in which two randomly selected entries are replaced with a small number $\delta_n < 0$. A sample simulated image along with mean, positive features, and negative features is shown in Figure 8.

In the experiments carried out in this scenario, δ_n is fixed at -0.3. We then experiment with different positive feature magnitudes δ_p. The recovery square root mean square error performance metrics for SSD (PP) and PFD are presented in Figure 9. As is clear from Figure 9, e_M for SSD (PP) is always higher than PFD and experiences a slight jump after $\delta_p = 0.9$, but it remains almost constant for PFD. Both methods have almost the same values for e_{A_p}. The performance metric e_{A_n} remains negligible for PFD, but it is always significant for the SSD (PP). This is due to the identifiability issue faced by SSD (PP) as discussed later. Combining the performance in the cases of positive and negative features, the metric clearly shows that PFD performs much better in comparison to SSD (PP). As can be seen in Figure 10, PFD maintains perfect precision and recall scores for both positive and negative features when δ_p is varied. Although SSD (PP) scores a perfect recall when δ_p varies from 0.3 to 0.9, it exhibits a fairly low precision because it carries many more unnecessary features along with the actual negative features. This is illustrated in Figures 11 and 12. Starting $\delta_p = 1.2$, SSD (PP) is unable to detect negative features. This is illustrated in Figures 13 and 14. Nick local thresholding recalls positive features but with very poor precision. It is also unable to recall negative features completely. Global thresholding is unable to recall positive and negative features with

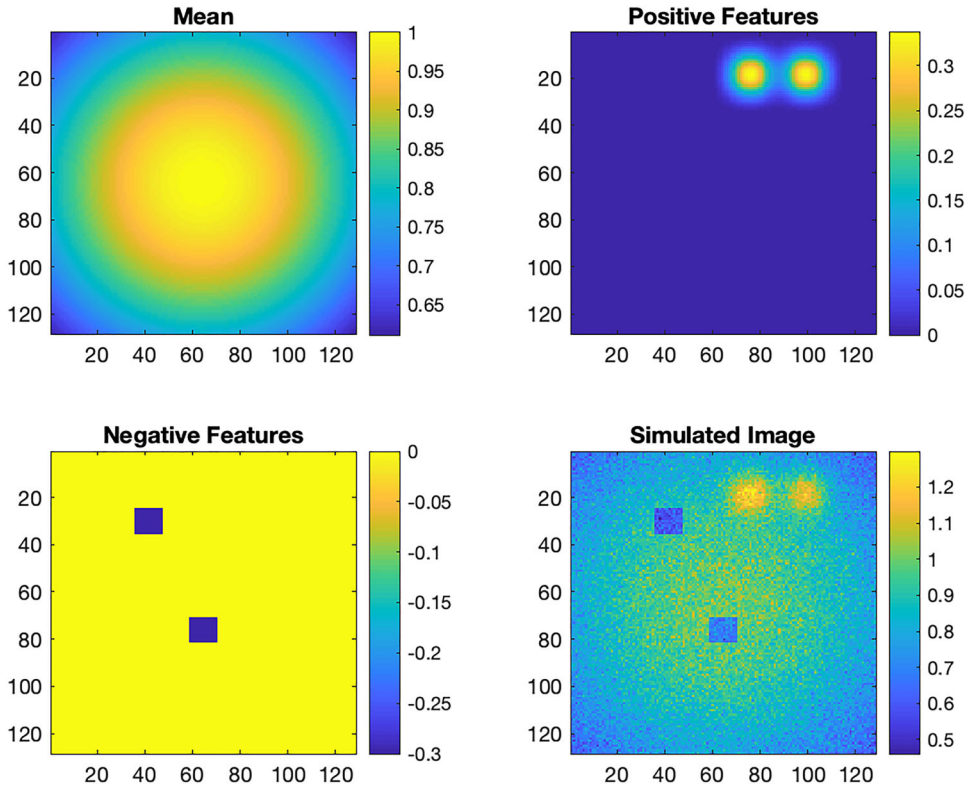

Figure 8. Scenario B – a sample simulated image with $\delta_p = 0.6$ and $\delta_n = -0.3$.

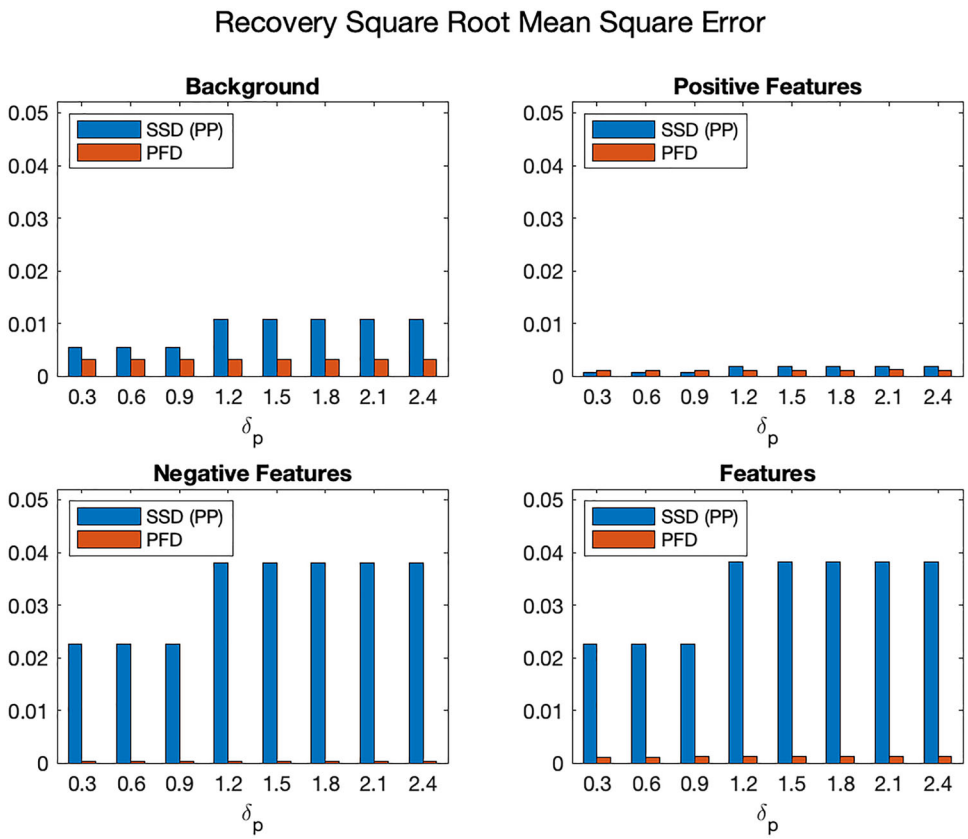

Figure 9. Scenario B – recovery square root mean square error results showing comparison between SSD (PP) and PFD.

precision. The sample visual results for Nick local and global thresholding are shown in Figure 15. In scenario B as well, PFD performs better than the existing methods. Overall, PFD outperforms benchmark methods in the simulation study.

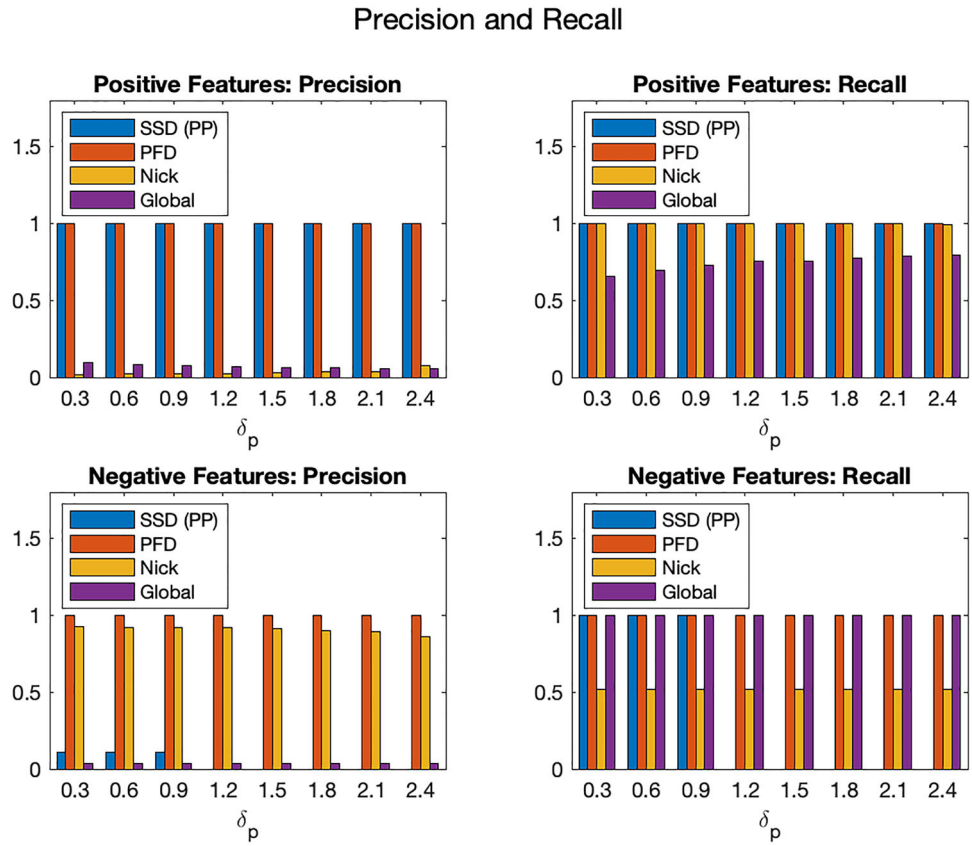

Figure 10. Scenario B – precision and recall results showing comparison among SSD (PP), PFD, Nick local thresholding, and global thresholding.

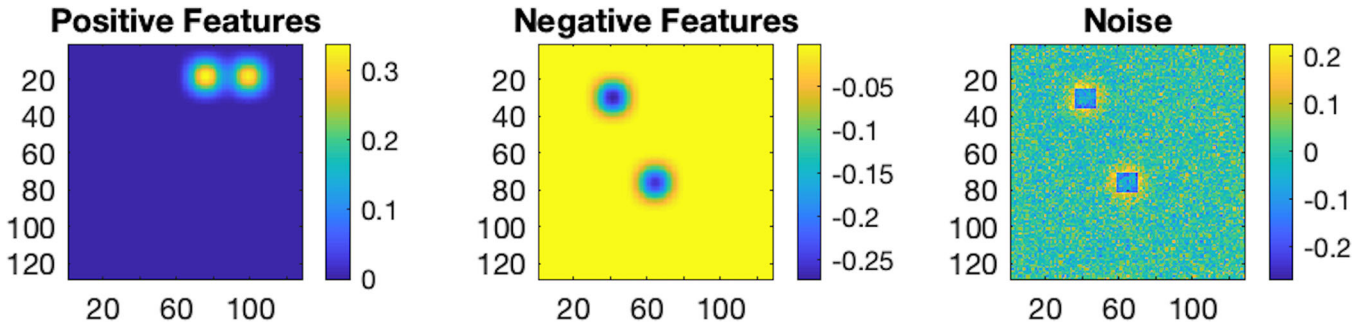

Figure 11. Scenario B – same basis issue faced by SSD (PP), when $\delta_p = 0.6$.

5. Case study

As discussed in Section I, CT image segmentation is an important task in the medical domain as it helps in computer-aided disease diagnosis and surgery planning. To illustrate the functionality of our proposed model and algorithm, we consider a two dimensional cross-sectional CT image of the human heart in the annulus region, as shown in Figure 1. This image has a special medical importance as it carries information on calcification and aortic root dimensions. We are interested in calcification (high-intensity region) detection and soft tissues (low-intensity segment) extraction. The proposed PFD model is employed to accomplish this task.

While applying the proposed PFD model to the aforementioned image with size 101×101, identity bases are used for detecting the calcification and the soft tissues (containing aortic valve). A cubic B-spline with 6×3 knots for the mean is selected. Hyperparameter λ is selected so that it minimizes the GCV as well as ensures convergence of the ADMM algorithm. Other tuning parameters (ρ, γ_p, γ_n) are dynamically tuned as discussed in Section III-B.2. To compare the performance, other benchmarks such as SSD with postprocessing, Otsu's global thresholding, and Nick local thresholding are also applied.

The results of different approaches are shown in Figure 16. As is clear from this figure, the proposed PFD model successfully extracts both the calcification (positive extreme features) and the soft tissues (negative extreme features) containing aortic valve structure. Next, we discuss post-processed results of the SSD. As far as calcification detection is concerned, there are too many false positives detected by this method. Moreover, the structure of the extracted soft tissues is not as crisp as the one detected by the PFD. Finally, we discuss results of the global thresholding and the

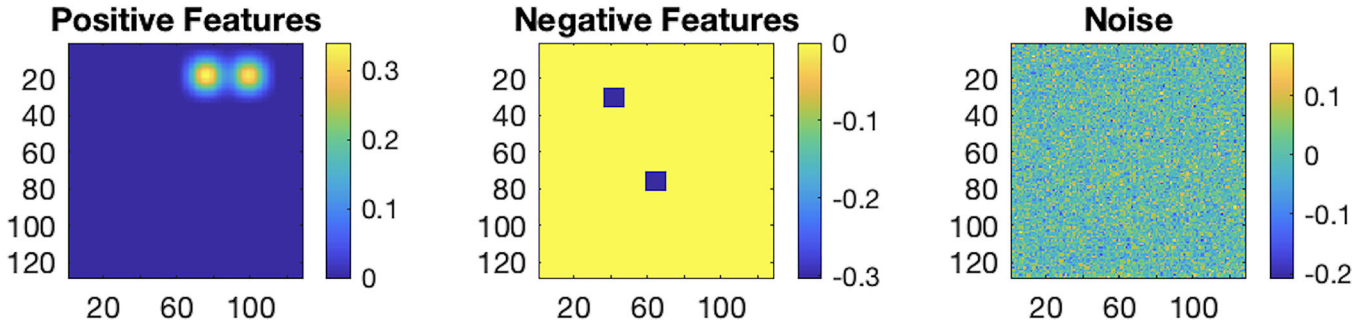

Figure 12. Scenario B – different bases successfully introduced using PFD, when $\delta_p = 0.6$.

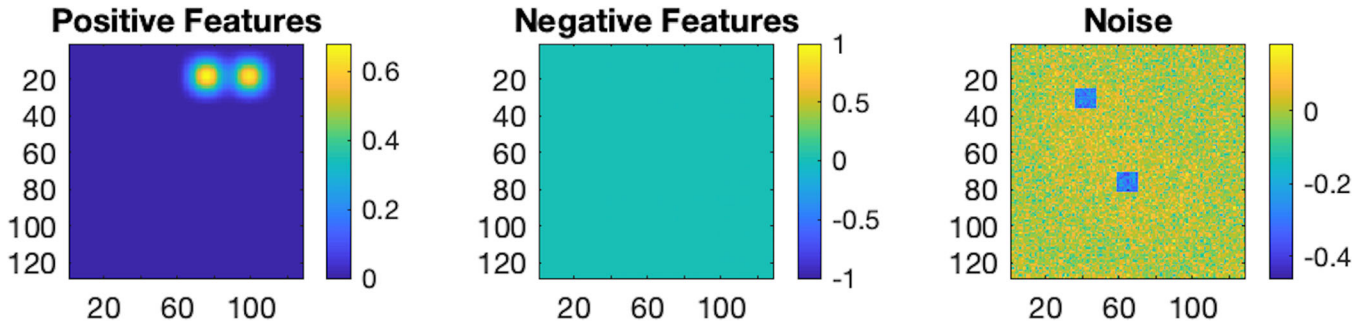

Figure 13. Scenario B – identifiability issue faced by SSD (PP), when $\delta_p = 1.2$.

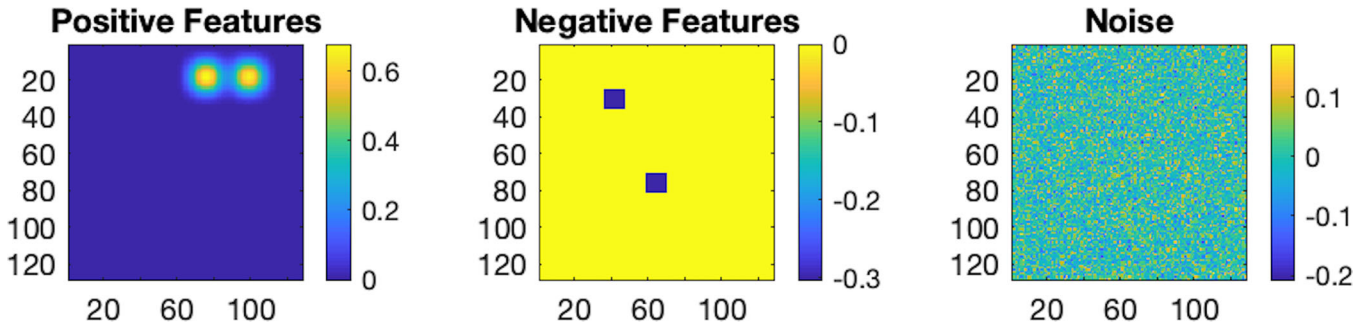

Figure 14. Scenario B – identifiability issue successfully dealt with PFD, when $\delta_p = 1.2$.

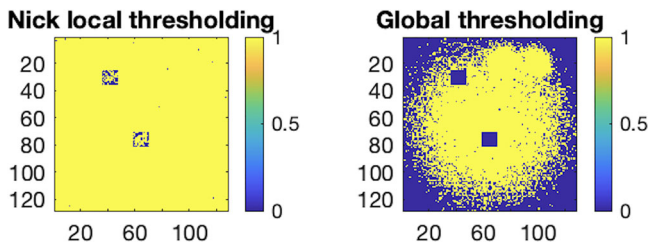

Figure 15. Scenario B – features detected by Nick local thresholding and global thresholding, when $\delta_p = 0.6$.

Nick thresholding. Since the global thresholding selects a single threshold based on intensity distribution, it clearly fails to extract and separate the calcification and the aortic root. The results of Nick thresholding mostly show the soft tissues region but that too with some false positives. Overall, the proposed PFD model outperforms the benchmarks.

The results obtained using the proposed PFD model have clear medical importance. Our model clearly identifies amount and location of the calcification. It helps physicians estimate the post-TAVR complication in terms of PVL. From the extracted soft tissues, aortic valve structure can easily be seen. This helps in selecting the appropriate size of the artificial valve to be implanted as a part of the TAVR surgery. This can again help in reducing the chances of post-TAVR complications in terms of PVL and stress induced in the aortic tissues. Furthermore, the proper segmentation of different features in the CT image can help develop a better digital model to be used as an input to the 3 D-printing which has shown the ability to assist physicians in TAVR surgery planning (Qian et al., 2017). Overall, the effectiveness of the proposed approach is quite convincing. Moreover, such an approach can be utilized for image segmentation tasks in other applications such as defect detection in manufacturing.

6. Conclusion

Computed tomography image segmentation is an important task in medicine. It is increasingly used for disease diagnosis and surgery planning. Aortic stenosis is one such heart disease where CT image is utilized for different purposes such

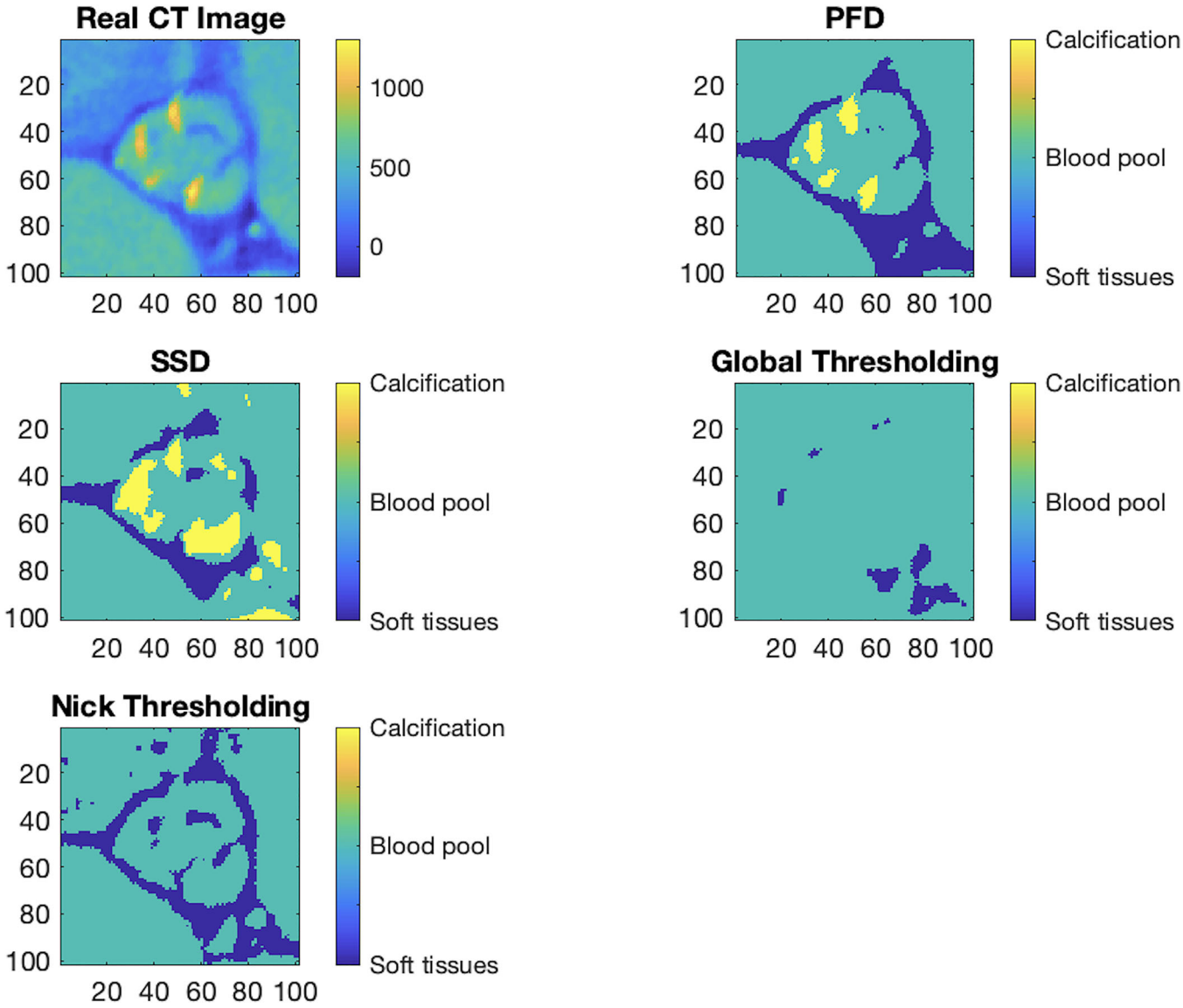

Figure 16. Calcification and soft tissues detection using PFD, SSD, global thresholding, and Nick local thresholding.

as initial diagnosis, clinical decision making, and surgery planning. Through the heart CT image, physicians explore the cause and extent of disease. For example, they try to locate calcification and its amount. Also, they are interested in understanding the structure of the aortic valve. With this knowledge, they are able to better understand the patient's condition and accordingly plan the treatment.

The major contribution of this work is to propose an approach to automatically quantify and locate the calcification, and extract the aortic valve, without the need for manual intervention. The proposed approach is flexible enough that it can be used for segmenting important features out of the CT image of other body parts as well. Our approach models image as consisting of four components – mean, positive extreme features, negative extreme features, and noise. The mean is further approximated using a set of smooth bases. Extreme features are also modeled using a set of appropriate bases. The approach treats mean as smooth and extreme features as sparse. To estimate the parameters of the proposed model, an optimization problem is formulated. This optimization problem is solved using a standard algorithm – alternating direction method of multipliers (ADMM). Within ADMM, an efficient proximal gradient method is utilized. Hyperparameter selection strategy is also discussed in detail. The performance of the proposed approach is tested using an extensive simulation study and a real-world case study. It is found that the PFD model outperforms benchmark methods in both the simulation study and the case study. In the case study, the results achieved by the proposed approach are not only precise but also of practical importance. The results are expected to play an important role in clinical decision making and surgery planning. Moreover, results can also act as an input for the digital model used in 3 D-printing technology for manufacturing virtual aortic valve.

The main focus of this work is to provide a framework for automatic detection of calcification and soft tissues in the 2-D CT image. This work opens up several directions

for future work. One possible extension is to apply this framework in a case of the 3-D CT image and try detecting the calcification and the aortic valve. Another interesting research idea is to extend this framework to a scenario where the CT image has a non-smooth mean. A new PFD model can be proposed where the non-smooth mean can be modeled using an appropriate basis. One suggestion for modeling the non-smooth mean can be the wavelet family (Huang et al., 2008; Yue et al., 2017). The appropriate wavelet type and level of decomposition can be selected on a case by case basis. Another interesting extension is to develop a data-driven approach to learn the appropriate basis to represent the features.

Acknowledgements

The authors would like to thank Jan Vlachy for his valuable comments and suggestions.

Disclosure statement

No potential conflict of interest was provided by the author(s).

ORCID

Geet Lahoti http://orcid.org/0000-0001-8876-4758
Jialei Chen http://orcid.org/0000-0002-6053-9302
Xiaowei Yue http://orcid.org/0000-0001-6019-0940
Hao Yan http://orcid.org/0000-0002-4322-7323
Chitta Ranjan http://orcid.org/0000-0003-1146-9325
Zhen Qian http://orcid.org/0000-0002-9867-0744
Chuck Zhang http://orcid.org/0000-0001-7681-5538
Ben Wang http://orcid.org/0000-0001-6484-1253

References

Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*, 641–647. https://doi.org/10.1109/34.295913

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends ® in Machine Learning*, *3*, 1–122. https://doi.org/10.1561/2200000016

Bradshaw, A. P., Taubman, D. S., Todd, M. J., Magnussen, J. S., & Halmagyi, G. M. (2013). Augmented active surface model for the recovery of small structures in ct. *IEEE Transactions on Image Processing*, *22*(11), 4394–4406. https://doi.org/10.1109/TIP.2013.2273666

Chen, J., Wang, K., Zhang, C., & Wang, B. (2018). An efficient statistical approach to design 3D-printed metamaterials for mimicking mechanical properties of soft biological tissues. *Additive Manufacturing*, *24*, 341–352. https://doi.org/10.1016/j.addma.2018.10.007

Chen, G., Xiang, D., Zhang, B., Tian, H., Yang, X., Shi, F., Zhu, W., Tian, B., & Chen, X. (2019). Automatic pathological lung segmentation in low-dose ct image using eigenspace sparse shape composition. *IEEE Transactions on Medical Imaging*, *38*(7), 1736–1749. https://doi.org/10.1109/TMI.2018.2890510

Chen, J., Xie, Y., Wang, K., Wang, Z. H., Lahoti, G., Zhang, C., Vannan, M. A., Wang, B., & Qian, Z. (2018). "*Generative Invertible Networks (GIN): Pathophysiology-Interpretable Feature Mapping and Virtual Patient Generation,*" In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol. 11070. Springer, pp. 537–545.

Chen, J., Xie, Y., Wang, K., Zhang, C., Vannan, M. A., Wang, B., & Qian, Z. "Active image synthesis for efficient labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2020).

Cootes, T., Taylor, C., Cooper, D., & Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, *61*, 38–59. https://doi.org/10.1006/cviu.1995.1004

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*, 407–499. https://doi.org/10.1214/009053604000000067

Eilers, P., & Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, *11*, 89–102. [Online]. Available: https://doi.org/10.1214/ss/1038425655

Gaines, B. R., Kim, J., & Zhou, H. (2018). Algorithms for fitting the constrained lasso. *Journal of Computational and Graphical Statistics*, *27*, 861–871. https://doi.org/10.1080/10618600.2018.1473777

Gao, X., Kitslaar, P. H., Scholte, A. J., Lelieveldt, B. P., Dijkstra, J., & J. H. Reiber. (2016). *Automatic aortic root segmentation in cta whole-body dataset*. In Medical Imaging 2016: Computer-Aided Diagnosis, vol. 9785. International Society for Optics and Photonics, 2016, p. 97850F. https://doi.org/10.1117/12.2216734

Gaw, N., Schwedt, T. J., Chong, C. D., Wu, T., & Li, J. (2018). A clinical decision support system using multi-modality imaging data for disease diagnosis. *IISE Transactions on Healthcare Systems Engineering*, *8*(1), 36–46. https://doi.org/10.1080/24725579.2017.1403520

Huang, J., Qian, Z., Huang, X., Metaxas, D., Axel, L. (2008). *"Tag Separation in Cardiac Tagged Mri," in International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 289–297.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, *1*(4), 321–331. https://doi.org/10.1007/BF00133570

Khurshid, K., Siddiqi, I., Faure, C., & N. Vincent. (2009). *Comparison of niblack inspired binarization methods for ancient documents*. In Document Recognition and Retrieval XVI, vol. 7247. International Society for Optics and Photonics, p. 72470U.

Kitslaar, P. H., Klooster, R. v., Staring, M., Lelieveldt, B. P., & R. J. van der Geest. (2015) *Segmentation of branching vascular structures using adaptive subdivision surface fitting*. In Medical Imaging 2015: Image Processing, vol. 9413. International Society for Optics and Photonics, p. 94133Z. https://doi.org/10.1117/12.2082222

Leymarie, F., & Levine, M. D. (1993). Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *15*(6), 617–634. https://doi.org/10.1109/34.216733

Nesterov, Y., & Nemirovskii, A. (1994). *Interior-Point Polynomial Algorithms in Convex Prgramming*. Society for Industrial and Applied Mathematics.

Nieuwenhuis, C., Toppe, E., & Cremers, D. (2013). A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *International Journal of Computer Vision*, *104*, 223–240. https://doi.org/10.1007/s11263-013-0619-y

Nosrati, M., & Hamarneh, G. (2016). "Incorporating prior knowledge in medical image segmentation: a survey," *arXiv:1607.01092*.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*, 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Parikh, N., & Boyd, S. (2013). Proximal algorithms. *Foundations and Trends in Optimization*, *1*, 123–231.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, *11*, 735–757. [Online]. Available: https://www.jstor.org/stable/1391159 https://doi.org/10.1198/106186002853

Sahoo, P., Soltani, S., & Wong, A. (1988). A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, *41*, 233–260. https://doi.org/10.1016/0734-189X(88)90022-9

Sanjay-Gopal, S., & Hebert, T. J. (1998). Bayesian pixel classification using spatially variant finite mixtures and the generalized em algorithm. *IEEE Transactions on Image Processing*, *7*(7), 1014–1028. https://doi.org/10.1109/83.701161

Sharma, N., & Aggarwal, L. (2010). Automated medical image segmentation techniques. *Journal of Medical Physics*, *35*, 3–14. https://doi.org/10.4103/0971-6203.58777

Sharma, N., & Aggarwal, L. M. (2010). Automated medical image segmentation techniques. *Journal of Medical Physics/Association of Medical Physicists of India*, *35*(1), 3. https://doi.org/10.4103/0971-6203.58777

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tseng, P. (2008). On accelerated proximal gradient methods for convex-concave optimization. *SIAM Journal on Optimization*, 1.

Tuncay, V., Prakken, N., van Ooijen, P., Budde, R., Leiner, T., & Oudkerk, M. (2015). Semiautomatic, quantitative measurement of aortic valve area using cta: Validation and comparison with transthoracic echocardiography. *BioMed Research International*, *2015* vol. https://doi.org/10.1155/2015/648283

Wang, Z. (2015). An efficient and robust method for automatically identifying the left ventricular boundary in cine magnetic resonance images. *IEEE Transactions on Automation Science and Engineering*, *13*(2), 536–542. https://doi.org/10.1109/TASE.2015.2403372

Wang, C., Komodakis, N., & Paragios, N. (2013). Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, *117*, 1610–1627. https://doi.org/10.1016/j.cviu.2013.07.004

Wang, Z. H., Lahoti, G., Wang, K., Liu, S., Zhang, C., Wang, B., Wu, C.-W., Vannan, M., & Z. Qian. (2018). Prediction of paravalvular leak post transcatheter aortic valve replacement using a convolutional neural network. In *2018 IEEE 15th International Symposium on Biomedical Imaging*. IEEE, pp. 1088–1091.

Xiao, L., Li, Y., & Ruppert, D. (2013). Fast bivariate p-splines: The sandwich smoother. *Journal of the Royal Statistical Society, Series B*, *75*, 577–599. [Online]. Available: https://doi.org/10.1111/rssb.12007

Xu, C., & Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, *7*(3), 359–369. https://doi.org/10.1109/83.661186

Yan, H., Paynabar, K., & Shi, J. (2017). Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, *59*, 102–114. https://doi.org/10.1080/00401706.2015.1102764

Yan, H., Paynabar, K., & Shi, J. (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, *60*(2), 181–197. https://doi.org/10.1080/00401706.2017.1346522

Yue, X., Yan, H., Park, J. G., Liang, Z., & Shi, J. (2017). A wavelet-based penalized mixed-effects decomposition for multichannel profile detection of in-line raman spectroscopy. *IEEE Transactions on Automation Science and Engineering*, *15*(3), 1258–1271. https://doi.org/10.1109/TASE.2017.2772218

Z., Qian, K. Wang, X., Zhou, V., Rajagopal, C., Meduri, Y., Chang, C., Wu, C., Zhang, B., Wang, & M. Vannan, (2017). Quantitative prediction of paravalvular leak in transcatheter aortic valve replacement based on tissue-mimicking 3d printing. *JACC Cardiovasc Imaging*, *10*, 719–731. https://doi.org/10.1016/j.jcmg.2017.04.005

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429. https://doi.org/10.1198/016214506000000735

Appendix I

Constrained weighted LASSO problem

The PFD problem in (3) is equivalent to a constrained weighted LASSO problem in the form of:

$$\begin{aligned} \underset{\theta_p,\theta_n}{\text{argmin}} \quad & (\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ & +\gamma_p\|\theta_p\|_1+\gamma_n\|\theta_n\|_1 \\ \text{subject to} \quad & \theta_p>0 \\ & \theta_n<0, \end{aligned} \tag{18}$$

where $H=B(B^TB+\lambda R)^{-1}B^T$.

Explanation: The PFD problem in (3) is first solved for θ while keeping θ_p & θ_n as fixed via following unconstrained optimization problem:

$$\text{argmin}_\theta \quad \|\tilde{y}-B\theta-B_p\theta_p-B_n\theta_n\|^2+\lambda\theta^TR\theta, \tag{19}$$

which gives, $\hat{\theta}=(B^TB+\lambda R)^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n)$. So, we can write $B\hat{\theta}=B(B^TB+\lambda R)^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n)=H(\tilde{y}-B_p\theta_p-B_n\theta_n)$. Next, we plug this into (3), we have the following:

$$\begin{aligned} \underset{\theta_p,\theta_n}{\text{argmin}} \quad & \|\tilde{y}-H(\tilde{y}-B_p\theta_p-B_n\theta_n)-B_p\theta_p-B_n\theta_n\|^2 \\ & +\lambda((B^TB+\lambda R)^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n))^TR \\ & (B^TB+\lambda R)^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ & +\gamma_p\|\theta_p\|_1+\gamma_n\|\theta_n\|_1 \\ \text{subject to} \quad & \theta_p>0 \\ & \theta_n<0. \end{aligned} \tag{20}$$

Next, we rewrite the first two terms in the objective function of the above problem (20) as follows:

$$\begin{aligned} &\|(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n)\|^2+\lambda(\tilde{y}-B_p\theta_p-B_n\theta_n)^T \\ &B(B^TB+\lambda R^T)^{-1}R(B^TB+\lambda R)^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n). \end{aligned}$$

Assuming R to be a symmetric matrix and denoting $K_\lambda=(B^TB+\lambda R)$, it again simplifies as follows:

$$\begin{aligned} &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I+H^2-2IH)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &\quad+\lambda(\tilde{y}-B_p\theta_p-B_n\theta_n)^TBK_\lambda^{-1}RK_\lambda^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I+BK_\lambda^{-1}B^TBK_\lambda^{-1}B^T-2BK_\lambda^{-1}B^T)(\tilde{y}-B_p\theta_p \\ &\quad-B_n\theta_n)+(\tilde{y}-B_p\theta_p-B_n\theta_n)^TBK_\lambda^{-1}\lambda RK_\lambda^{-1}B^T(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I+BK_\lambda^{-1}B^TBK_\lambda^{-1}B^T-2BK_\lambda^{-1}B^T \\ &\quad+BK_\lambda^{-1}\lambda RK_\lambda^{-1}B^T)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I+BK_\lambda^{-1}(B^TB+\lambda R)K_\lambda^{-1}B^T \\ &\quad-2BK_\lambda^{-1}B^T)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I+BK_\lambda^{-1}K_\lambda K_\lambda^{-1}B^T-2BK_\lambda^{-1}B^T)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-BK_\lambda^{-1}B^T)(\tilde{y}-B_p\theta_p-B_n\theta_n) \\ &=(\tilde{y}-B_p\theta_p-B_n\theta_n)^T(I-H)(\tilde{y}-B_p\theta_p-B_n\theta_n). \end{aligned}$$

Substituting this result into 20, we get the revised optimization problem for estimating θ_p & θ_n as claimed in 18. During the above simplification process, we discover that following relation holds true:

$$(I-H)=(I-H)^T(I-H)+BK_\lambda^{-1}\lambda RK_\lambda^{-1}B^T. \tag{21}$$

It is to be noted that a relationship similar to (21) is also claimed to be true in Proposition 3 of (Yan et al., 2017). We have given the detailed proof for the sake of completeness and reader's understanding.

Appendix II

PG method assumption

Proof of Proposition 1: Gradient of $F(\theta_p)$ can be calculated as shown below:

$$\begin{aligned} \nabla F(\theta_p)=&\,2B_p^T(I-H)(B_p\theta_p+B_n\theta_n^{(k-1)}-\tilde{y}) \\ &+\rho(\theta_p-z_p^{(k-1)}+u_p^{(k-1)}). \end{aligned} \tag{22}$$

Note that $\|X\|_2$ represents spectral norm of matrix X. Next, we calculate the following quantity:

$$\begin{aligned} \|\nabla F(\alpha)-\nabla F(\beta)\| &= \|2B_p^T(I-H)B_p(\alpha-\beta)+\rho(\alpha-\beta)\| \\ &\le \|2B_p^T(I-H)B_p(\alpha-\beta)\|+\rho\cdot\|(\alpha-\beta)\| \\ &\le \|2B_p^T(I-H)B_p\|_2\cdot\|(\alpha-\beta)\|+\rho\cdot\|(\alpha-\beta)\| \\ &\le \|2B_p^T\|_2\cdot\|(I-H)\|_2\cdot\|B_p\|_2\cdot\|(\alpha-\beta)\| \\ &\quad+\rho\cdot\|(\alpha-\beta)\|. \end{aligned} \tag{23}$$

Using the result $\|I-H\|_2 \leq 1$ from claim 5 in (Yan et al., 2017), it further simplifies as follows:

$$\begin{aligned}&\|\nabla F(\alpha)-\nabla F(\beta)\|\\&\leq \|2B_p^T\|_2\cdot\|B_p\|_2\cdot\|(\alpha-\beta)\|+\rho\cdot\|(\alpha-\beta)\|\\&\leq (2\|B_p\|_2^2+\rho)\|(\alpha-\beta)\|.\end{aligned}\tag{24}$$

It is easy to deduce that $L=2\|B_p\|_2^2+\rho$.

Appendix III
PG method closed-form solution

Proof of Proposition 2: We are required to solve the following problem:

$$\begin{aligned}\underset{\theta_p}{\text{argmin}}\quad & F(\theta_p^{(k-1)})+\langle\theta_p-\theta_p^{(k-1)},\nabla F(\theta_p^{(k-1)})\rangle\\&+\frac{L}{2}\|\theta_p-\theta_p^{(k-1)}\|^2+\gamma_p\|\theta_p\|_1.\end{aligned}\tag{25}$$

First, we simplify the objective function in the following fashion:

$$\begin{aligned}&F(\theta_p^{(k-1)})+\langle\theta_p-\theta_p^{(k-1)},\nabla F(\theta_p^{(k-1)})\rangle+\frac{L}{2}\|\theta_p-\theta_p^{(k-1)}\|^2+\gamma_p\|\theta_p\|_1\\&=(\theta_p-\theta_p^{(k-1)})^T(2B_p^T(I-H)(B_p\theta_p^{(k-1)}+B_n\theta_n^{(k-1)}-\tilde{y})\\&\quad+\rho(\theta_p^{(k-1)}-z_p^{(k-1)}+u_p^{(k-1)}))\\&\quad+\frac{L}{2}\|\theta_p-\theta_p^{(k-1)}\|^2+\gamma_p\|\theta_p\|_1.\end{aligned}$$

Multiplying by $\frac{2}{L}$ and adding an appropriate constant term to make a perfect square, it again simplifies as follows:

$$\begin{aligned}&=\|\theta_p-\theta_p^{(k-1)}+\frac{2}{L}B_p^T(B_p\theta_p^{(k-1)}+B_n\theta_n^{(k-1)}-\tilde{y}+H(\tilde{y}-B_p\theta_p^{(k-1)}\\&\quad-B_n\theta_n^{(k-1)})+\rho(\theta_p^{(k-1)}-z_p^{(k-1)}+u_p^{(k-1)})\|^2+\frac{2}{L}\gamma_p\|\theta_p\|_1\\&=\|\theta_p-\theta_p^{(k-1)}-\frac{2}{L}B_p^T(\tilde{y}-B\theta^{(t-1)}-B_p\theta_p^{(k-1)}-B_n\theta_n^{(k-1)})\\&\quad+\rho(\theta_p^{(k-1)}-z_p^{(k-1)}+u_p^{(k-1)})\|^2+\frac{2}{L}\gamma_p\|\theta_p\|_1.\end{aligned}$$

Substituting this in (25) and we get the revised form of the problem as shown below:

$$\begin{aligned}\underset{\theta_p}{\text{argmin}}\quad & \|\theta_p-\theta_p^{(k-1)}-\frac{2}{L}B_p^T(\tilde{y}-B\theta^{(t-1)}-B_p\theta_p^{(k-1)}\\&-B_n\theta_n^{(k-1)})+\rho(\theta_p^{(k-1)}-z_p^{(k-1)}+u_p^{(k-1)})\|^2\\&+\frac{2}{L}\gamma_p\|\theta_p\|_1.\end{aligned}\tag{26}$$

Next, it is easy to derive that the solution to the problem in 26 is given by the following soft-thresholding operator:

$$\begin{aligned}\theta_p^{(k)}=S_{\frac{\gamma_p}{L}}(&\theta_p^{(k-1)}+\frac{2}{L}B_p^T(\tilde{y}-B\theta^{(t-1)}-B_p\theta_p^{(k-1)}\\&-B_n\theta_n^{(k-1)})-\rho(\theta_p^{(k-1)}-z_p^{(k-1)}+u_p^{(k-1)})).\end{aligned}\tag{27}$$