Generalized nonconvex hyperspectral anomaly detection via background representation learning with dictionary constraint

Quan Yu (HNU & SRIBD)

Joint work with Minru Bai

第一届全国算法软件与应用研讨会

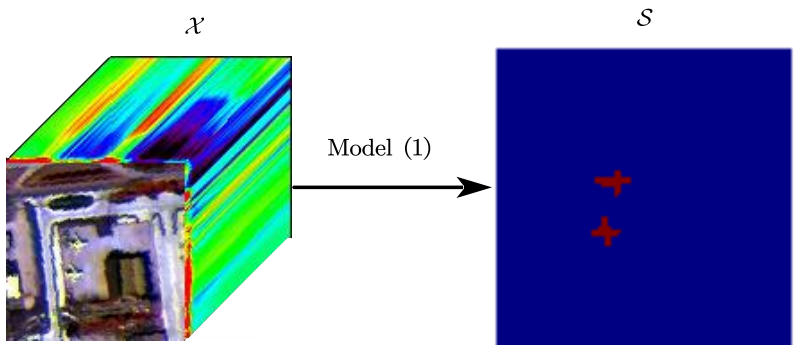
Contents

Contents

TLRR model

Given an observed Hyperspectral image (HSI) $\mathcal{X}$ and a dictionary $\mathcal{A}$, the tensor low rank representation (TLRR) model for hyperspectral anomaly detection (HAD) can be expressed as

$$\min_{\mathcal{L},\mathcal{S}} \operatorname{rank}(\mathcal{L}) + \lambda \|\mathcal{S}\|_{\ell_{F,0}}{}^{1}, \quad \text{s.t.} \quad \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}. \tag{1}$$



[1] $\|\mathcal{S}\|_{\ell_{F,0}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|\mathcal{S}(i,j,:)\|_F^0$

- Construction of dictionary
  - SVD: TLRA-MSL[2]
  - k-means clustering: GTVLRR[3], SRTDaAW[4]
  - RPCA: TLRSR[5]
- Low rank approximation
  - Nuclear norm: GTVLRR[3], LRASR[6]
  - Weighted nuclear norm: TLRSR[5]
  - Truncated nuclear norm: PTA[7]

---

[2] He, Xu, et al. "Anomaly Detection for Hyperspectral Imagery via Tensor Low-Rank Approximation With Multiple Subspace Learning." IEEE Transactions on Geoscience and Remote Sensing (2023).

[3] Cheng, Tongkai, and Bin Wang. "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection." IEEE Transactions on Geoscience and Remote Sensing 58.1 (2019): 391-406.

[4] Yang, Yixin, et al. "Hyperspectral anomaly detection through sparse representation with tensor decomposition-based dictionary construction and adaptive weighting." IEEE Access 8 (2020): 72121-72137.

[5] Wang, Minghua, et al. "Learning tensor low-rank representation for hyperspectral anomaly detection." IEEE Transactions on Cybernetics 53.1 (2022): 679-691.

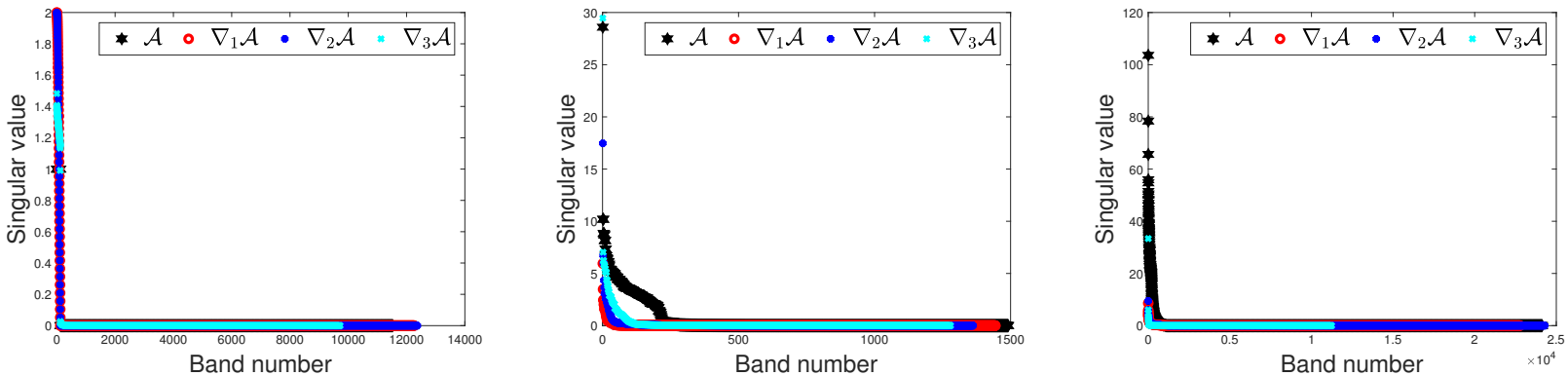
[6] Xu, Yang, et al. "Anomaly detection in hyperspectral images based on low-rank and sparse representation." IEEE Transactions on Geoscience and Remote Sensing 54.4 (2015): 1990-2000.

[7] Li, Lu, et al. "Prior-based tensor approximation for anomaly detection in hyperspectral imagery." IEEE Transactions on Neural Networks and Learning Systems 33.3 (2020): 1037-1050.

Contents

Dictionary constraint

Different from model (1) which simply adopts a pre-built dictionary $\mathcal{A}$, we perform dictionary construction and anomaly detection simultaneously to learn a more comprehensive dictionary for background reconstruction.



(a) SVD (b) k-means clustering (c) RPCA

Figure 1: Singular values of dictionary tensor $\mathcal{A}$ and gradient tensor $\nabla_u\mathcal{A}$, $u \in [3]$.

We exploit this property to model the dictionary tensors using the gradient tensors, and it can be modeled as follows:

$$\begin{aligned} \min_{\mathcal{A},\mathcal{L},\mathcal{S}} \ & \sum_{u=1}^{3} \alpha_u \operatorname{rank}\left(\nabla_u \mathcal{A}\right) + \lambda_1 \operatorname{rank}\left(\mathcal{L}\right) + \lambda_2 \left\|\mathcal{S}\right\|_{\ell_{F,0}} \\ \text{s.t.} \ & \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}, \end{aligned} \tag{2}$$

where $\nabla_1\mathcal{A}$, $\nabla_2\mathcal{A}$, $\nabla_3\mathcal{A}$ denote the first-order forward finite-difference operators along the vertical, horizontal, and spectral directions, respectively.

Generalized nonconvex approximation

Different from most existing nonconvex HAD methods that only use nonconvex approximation for low rank, we apply it to both low rank and sparsity. Moreover, we pursue a general surrogate for approximating low rank and sparsity, i.e.,

$$\begin{aligned} \|\mathcal{X}\|_{\psi} &= \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{\min\{n_1,n_2\}}\psi\left(\sigma_i\left(\bar{X}^{(k)}\right)\right), \\ \|\mathcal{X}\|_{\ell_{F,1}^{\psi}} &= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\psi\left(\|\mathcal{X}(i,j,:)\|_F\right), \end{aligned} \tag{3}$$

where $\psi(\cdot):\mathbb{R}_+\to\mathbb{R}_+$ is a function.

By replacing the $\operatorname{rank}(\cdot)$ and $\|\cdot\|_{\ell_{F,0}}$ with (3), our GNBRL model is formulated as:

$$\begin{aligned} \min_{\mathcal{A},\mathcal{L},\mathcal{S}} \ & \sum_{u=1}^{3} \alpha_u \|\nabla_u \mathcal{A}\|_{\psi} + \lambda_1 \|\mathcal{L}\|_{\psi} + \lambda_2 \|\mathcal{S}\|_{\ell_{F,1}^{\psi}} \\ \text{s.t.} \ & \mathcal{X} = \mathcal{A} * \mathcal{L} + \mathcal{S}. \end{aligned} \tag{4}$$

Assumption 3.1

The function $\psi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies: ψ is continuous, nondecreasing and concave with $\psi(0) = 0$.

定理 1

Most functions satisfy Assumption 3.1. We list five of them as follows. Here we only consider the case $x > 0$.

(1) *L1:* $\psi^{L1}(x) = x$;

(2) *Lp:* $\psi^{Lp}(x) = x^p$, $p \in (0,1)$;

(3) *MCP:* $\psi^{MCP}(x) = \begin{cases} x - \frac{x^2}{2\alpha}, & 0 \le x \le \alpha, \\ \frac{\alpha}{2}, & x > \alpha \end{cases}$ *with* $\alpha > 0$;

(4) *Logarithm:* $\psi^{Log}(x) = \log(\frac{x}{\theta} + 1)$ *with* $\theta > 0$;

(5) *Capped folded functions:*

- *Capped* $L1$: $\psi^{CapL1}(x) = \min\left\{1, \frac{x}{v}\right\}$;
- *Capped* Lp : $\psi^{CapLp}(x) = \min\left\{1, \frac{x^p}{v^p}\right\}$, $p \in (0,1)$;
- *Capped MCP:*
  $\psi^{CapMCP}(x) = \min\left\{1, \frac{2\alpha}{\nu(2\alpha-\nu)}\psi^{MCP}(x)\right\}$, $0 < \nu < \alpha$;
- *Capped Logarithm:* $\psi^{CapLog}(x) = \min\left\{1, \frac{1}{\psi^{Log}(v)}\psi^{Log}(x)\right\}$.

CF2 framework for GNBRL

To improve the detection accuracy of complex objects in a cluttered scene, we develop a coarse to fine two-stage (CF2) framework for GNBRL (CF2-GNBRL).

1) Coarse Stage: In the coarse stage, a coarse anomaly $\tilde{\mathcal{S}}$ is obtained by applying the GNBRL model to the whole HSI.
2) Fine Stage: We first divide the whole HSI into N patches third order sub-tensors according to BM3D. Then we apply the GNBRL model to each sub-tensor to obtain $\hat{\mathcal{S}}^1_{patch}, \hat{\mathcal{S}}^2_{patch}, \ldots, \hat{\mathcal{S}}^N_{patch}$.
Next, we divide $\tilde{\mathcal{S}}$ into N patches following the partitions employed in the current fine stage to obtain $\tilde{\mathcal{S}}^1_{patch}, \tilde{\mathcal{S}}^2_{patch}, \ldots, \tilde{\mathcal{S}}^N_{patch}$.

Finally, we obtain $\mathcal{S}^{\star}$ by

$$\mathcal{S}^{\star,l}_{patch} = \begin{cases} \tilde{\mathcal{S}}^{l}_{patch}, & \text{if } \operatorname{gap}\left(\tilde{\mathcal{S}}^{l}_{patch}, \hat{\mathcal{S}}^{l}_{patch}\right) < \varrho, \\ \hat{\mathcal{S}}^{l}_{patch}, & \text{if } \operatorname{gap}\left(\tilde{\mathcal{S}}^{l}_{patch}, \hat{\mathcal{S}}^{l}_{patch}\right) \geq \varrho, \end{cases} \tag{5}$$

where ϱ is a given parameter and

$$\operatorname{gap}\left(\tilde{\mathcal{S}}^{l}_{patch}, \hat{\mathcal{S}}^{l}_{patch}\right) = \frac{\left\|\tilde{\mathcal{S}}^{l}_{patch} - \hat{\mathcal{S}}^{l}_{patch}\right\|_F}{\left\|\tilde{\mathcal{S}}^{l}_{patch}\right\|_F}.$$

Error bound of GNBRL

In the following, we present some properties of the nonconvex function ψ, which are essential for the error bound analysis.

定理 2

Suppose that $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are two arbitrary tensors. Then, the following properties hold.

(1) $\|\mathcal{B} - \mathcal{S}\|_\psi \geq \|\mathcal{B}\|_\psi - \|\mathcal{S}\|_\psi$;

(2) $\|\mathcal{B} - \mathcal{S}\|_{\ell^\psi_{F,1}} \leq \|\mathcal{B}\|_{\ell^\psi_{F,1}} + \|\mathcal{S}\|_{\ell^\psi_{F,1}}$;

(3) $\psi\left(\|\mathcal{B}\|_F\right) \leq \|\mathcal{B}\|_{\ell^\psi_{F,1}} \leq \|\mathcal{B}\|_{\psi,1}$, *where* $\|\mathcal{B}\|_{\psi,1} := \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\psi\left(|\mathcal{B}_{ijk}|\right)$.

定理 3

Let $\left(\mathcal{L}^\natural, \mathcal{S}^\natural\right)$ *be the pair of true low rank and sparse tensors, and* $(\mathcal{A}^\star, \mathcal{L}^\star, \mathcal{S}^\star)$ *be an optimal solution to the optimization problem* (4). *Assume that* $\mathcal{A}^\star$ *satisfies* ψ*-RTEC(s),* $\mathcal{X} = \mathcal{A}^\star * \mathcal{L}^\natural + \mathcal{S}^\natural$, $\left\|\mathcal{L}^\natural\right\|_\psi \leq \left\|\mathcal{L}^\star\right\|_\psi := s$, *and* $\lambda_2 > \lambda_1 r \vartheta^\psi_{r,s}$ *with* $r = \min\{n_1, n_2\}$. *Then we have*

$$\psi\left(\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F\right) \leq \left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_{\ell^\psi_{F,1}} \leq \frac{2\lambda_2 \left\|\mathcal{S}^\natural\right\|_{\ell^\psi_{F,1}}}{\lambda_2 - \lambda_1 r \vartheta^\psi_{r,s}}, \tag{6}$$

where $\vartheta^\psi_{r,s}$ *is a constant that depends on* r, s, ψ.

推论 4

When $\psi(x) = x^p$, the average of the entries of the sparse component $\mathcal{S}^\natural$ is bounded by T and the cardinality of the support $\mathcal{S}^\natural$ is bounded by m. By properly choosing λ_1 and λ_2, we have

$$\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F \leq \sqrt[p]{4mT}.$$

That is, for very sparse anomaly tensors, as long as T is bounded, then $\left\|\mathcal{S}^\natural - \mathcal{S}^\star\right\|_F / M$ with $M = n_1 n_2 n_3$ is rather small, indicating good recovery.

Contents

In real-world HSIs data, the entries in a spectral vector are corrupted by Gaussian noise, so we convert (4) to the following problem:

$$\min_{\mathcal{A},\mathcal{L},\mathcal{S}} \sum_{u=1}^{3} \alpha_u \left\|\nabla_u \mathcal{A}\right\|_{\psi} + \lambda_1 \left\|\mathcal{L}\right\|_{\psi} + \lambda_2 \left\|\mathcal{S}\right\|_{\ell_{F,1}^{\psi}} + \beta f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right), \quad (7)$$

where $f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right) = \frac{1}{2}\left\|\mathcal{A} * \mathcal{L} + \mathcal{S} - \mathcal{X}\right\|_F^2$.

Now we develop an ELADMM algorithm to solve problem (7). By introducing the auxiliary variable $\mathcal{C}_u = \nabla_u \mathcal{A}$, $u \in [3]$, problem (7) can be rewritten as:

$$\begin{aligned} \min_{\mathcal{A}, \mathcal{L}, \mathcal{S}, \{\mathcal{C}_u\}_{u=1}^3} \quad & \sum_{u=1}^{3} \alpha_u \left\|\mathcal{C}_u\right\|_\psi + \lambda_1 \left\|\mathcal{L}\right\|_\psi + \lambda_2 \left\|\mathcal{S}\right\|_{\ell_{F,1}^\psi} + \beta f\left(\mathcal{A}, \mathcal{L}, \mathcal{S}\right), \\ \text{s.t.} \quad & \mathcal{C}_u = \nabla_u \mathcal{A},\ u \in [3]. \end{aligned} \tag{8}$$

The augmented Lagrangian function of (8) can be given by

$$
\begin{aligned}
& L\left(\mathcal{S}, \mathcal{A}, \mathcal{C}_u, \mathcal{L}; \mathcal{T}_u, \beta_u\right) \\
= & \sum_{u=1}^{3}\left(\alpha_u\left\|\mathcal{C}_u\right\|_\psi+\left\langle\mathcal{T}_u, \nabla_u \mathcal{A}-\mathcal{C}_u\right\rangle+\frac{\beta_u}{2}\left\|\nabla_u \mathcal{A}-\mathcal{C}_u\right\|_F^2\right) \\
& +\lambda_1\|\mathcal{L}\|_\psi+\lambda_2\|\mathcal{S}\|_{\ell_{F,1}^\psi}+\beta f(\mathcal{A}, \mathcal{L}, \mathcal{S}),
\end{aligned}
\tag{9}
$$

where β_u for $u \in [3]$ are the penalty parameter, and $\mathcal{T}_u$ for $u \in [3]$ are the Lagrange multipliers.

Under the ADMM algorithm framework, we can alternatively update each variable:

$$\begin{cases} \mathcal{S}^{t+1} \in \operatorname{argmin}_{\mathcal{S}} \ L\left(\mathcal{S}, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right), \\ \mathcal{A}^{t+1} \in \operatorname{argmin}_{\mathcal{A}} \ L\left(\mathcal{S}^{t+1}, \mathcal{A}, \mathcal{C}_u^t, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right), \\ \mathcal{C}_u^{t+1} \in \operatorname{argmin}_{\mathcal{C}_u} \ L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u, \mathcal{L}^t; \mathcal{T}_u^t, \beta_u^t\right), \\ \mathcal{L}^{t+1} \in \operatorname{argmin}_{\mathcal{L}} \ L\left(\mathcal{S}^{t+1}, \mathcal{A}^{t+1}, \mathcal{C}_u^{t+1}, \mathcal{L}; \mathcal{T}_u^t, \beta_u^t\right), \\ \mathcal{T}_u^{t+1} = \mathcal{T}_u^t + \beta_u^t \left(\nabla_u \mathcal{A}^{t+1} - \mathcal{C}_u^{t+1}\right), \ \beta_u^{t+1} = \rho \beta_u^t, \end{cases}$$

where t denotes the iteration number and ρ is a constant value greater than 1.

Although $L\left(\mathcal{S},\mathcal{A},\mathcal{C}_u,\mathcal{L};\mathcal{T}_u,\beta_u\right)$ with respect to $\mathcal{A}$ is convex, there is no closed-form solution. We update $\mathcal{A}$ by solving the following sub-problem:

$$\begin{aligned}\arg\min_{\mathcal{A}} \sum_{u=1}^{3}\left(\left\langle \mathcal{T}_u^t, \nabla_u\mathcal{A}-\mathcal{C}_u^t\right\rangle+\frac{\beta_u^t}{2}\left\|\nabla_u\mathcal{A}-\mathcal{C}_u^t\right\|_F^2\right)\\ +\beta\left\langle\nabla_{\mathcal{A}} f\left(\hat{\mathcal{A}}^t,\mathcal{L}^t,\mathcal{S}^{t+1}\right),\mathcal{A}-\hat{\mathcal{A}}^t\right\rangle+\frac{\beta l_{\mathcal{A}}^t}{2}\left\|\mathcal{A}-\hat{\mathcal{A}}^t\right\|_F^2,\end{aligned} \tag{10}$$

where $f\left(\mathcal{A},\mathcal{L},\mathcal{S}\right)=\frac{1}{2}\left\|\mathcal{A}*\mathcal{L}+\mathcal{S}-\mathcal{X}\right\|_F^2$. $l_{\mathcal{A}}^t\geq l_{\mathcal{A}}(f)$, $l_{\mathcal{A}}(f)$ is a Lipschitz constant of $\nabla_{\mathcal{A}} f\left(\mathcal{A},\mathcal{L}^t,\mathcal{S}^{t+1}\right)$ with respect to $\mathcal{A}$, and $\hat{\mathcal{A}}^t$ is an extrapolated point.

Algorithm 1: ELADMM method to solve GNBRL

Input: The tensor data $\mathcal{X}$, parameters $\{\alpha_u\}_{u=1}^3$, λ_1, λ_2, β.

Initialize: $\mathcal{A}^0$, $\mathcal{L}^0$, $\mathcal{S}^0$, $\{\mathcal{C}_u^0, \mathcal{T}_u^0\}_{u=1}^3$, $\{\beta_u^0\}_{u=1}^3$.

While not converge do

Step **1**. Update $\mathcal{S}^{t+1}$.

Step **2**. Let $\hat{\mathcal{A}}^t = \mathcal{A}^t + \omega_{\mathcal{A}}^t \left(\mathcal{A}^t - \mathcal{A}^{t-1}\right)$.

Step **3**. Update $\mathcal{A}^{t+1}$.

Step **4**. Update $\mathcal{C}_u^{t+1}$.

Step **5**. Let $\hat{\mathcal{L}}^t = \mathcal{L}^t + \omega_{\mathcal{L}}^t \left(\mathcal{L}^t - \mathcal{L}^{t-1}\right)$.

Step **6**. Update $\mathcal{L}^{t+1}$.

Step **7**. Update multipliers $\mathcal{T}_u^{t+1}$ and penalty parameters β_u^{t+1}.

Let $t := t + 1$ and go to Step **1**.

end while

Output: $\mathcal{S}^{t+1}$, $\mathcal{A}^{t+1}$, $\mathcal{L}^{t+1}$.

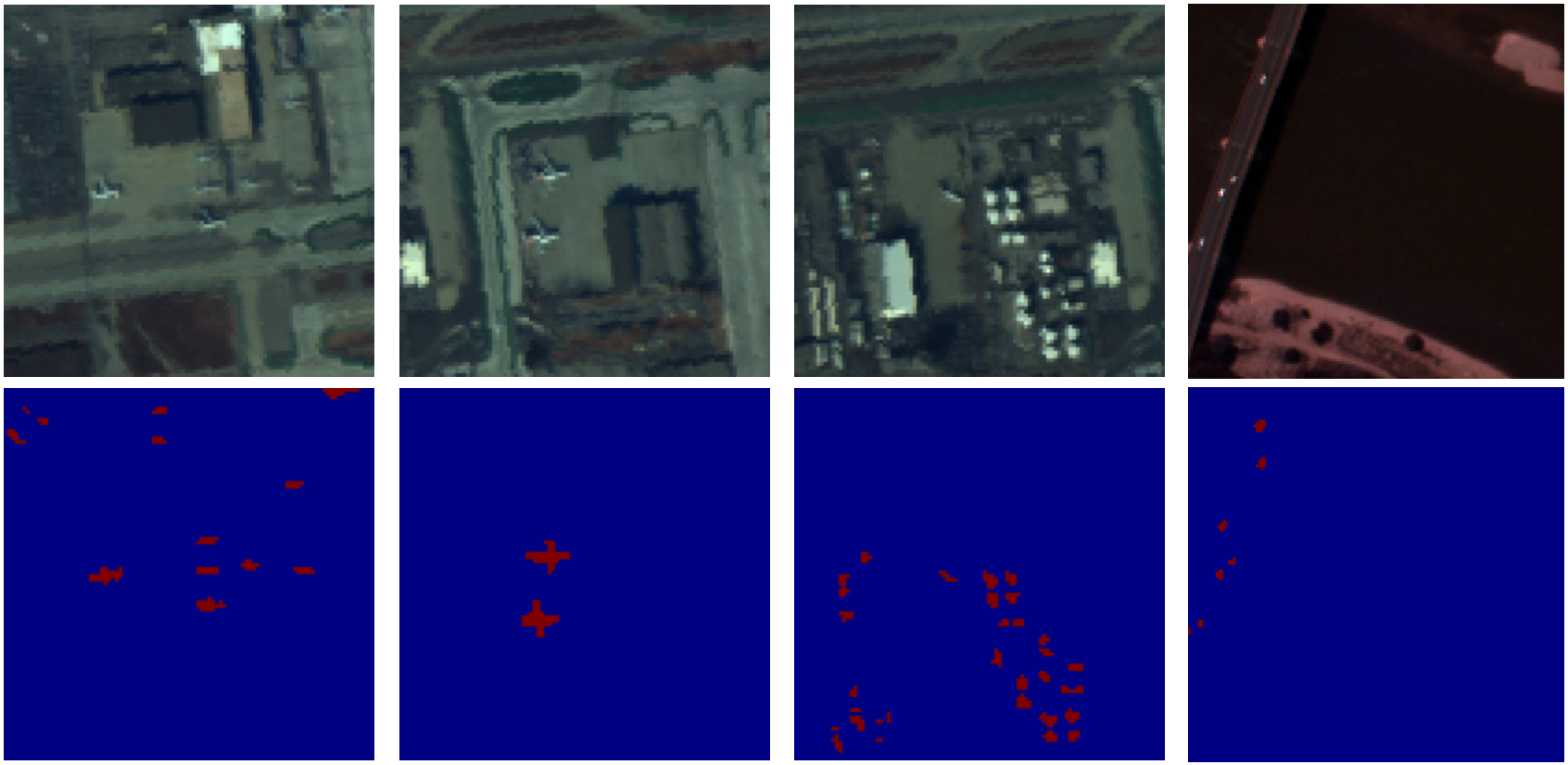
Convergence analysis

定理 5

Let $\{\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t, \mathcal{T}_u^t\}$ *be a sequence generated by Algorithm 1. Suppose that the sequence* $\{\mathcal{A}^t, \mathcal{L}^t\}_{t=1}^{\infty}$ *is bound. Then any accumulation point of the sequence* $\{\mathcal{S}^t, \mathcal{A}^t, \mathcal{C}_u^t, \mathcal{L}^t, \mathcal{T}_u^t\}$ *is a Karush-Kuhn-Tucker (KKT) point of the optimization problem* (8).

Contents

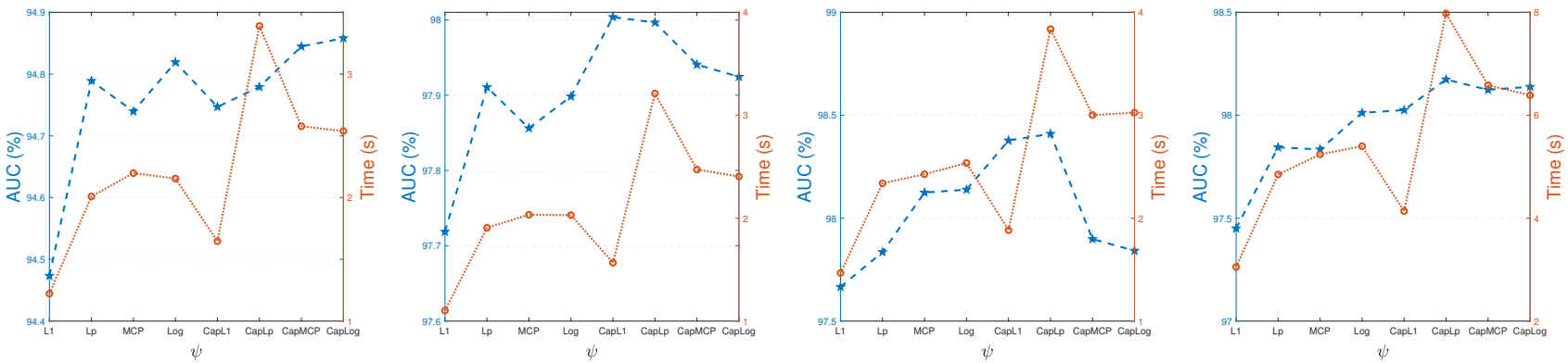
Hyperspectral data



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 2: Pseudo-color images and ground-truth maps of the four HSIs data sets.

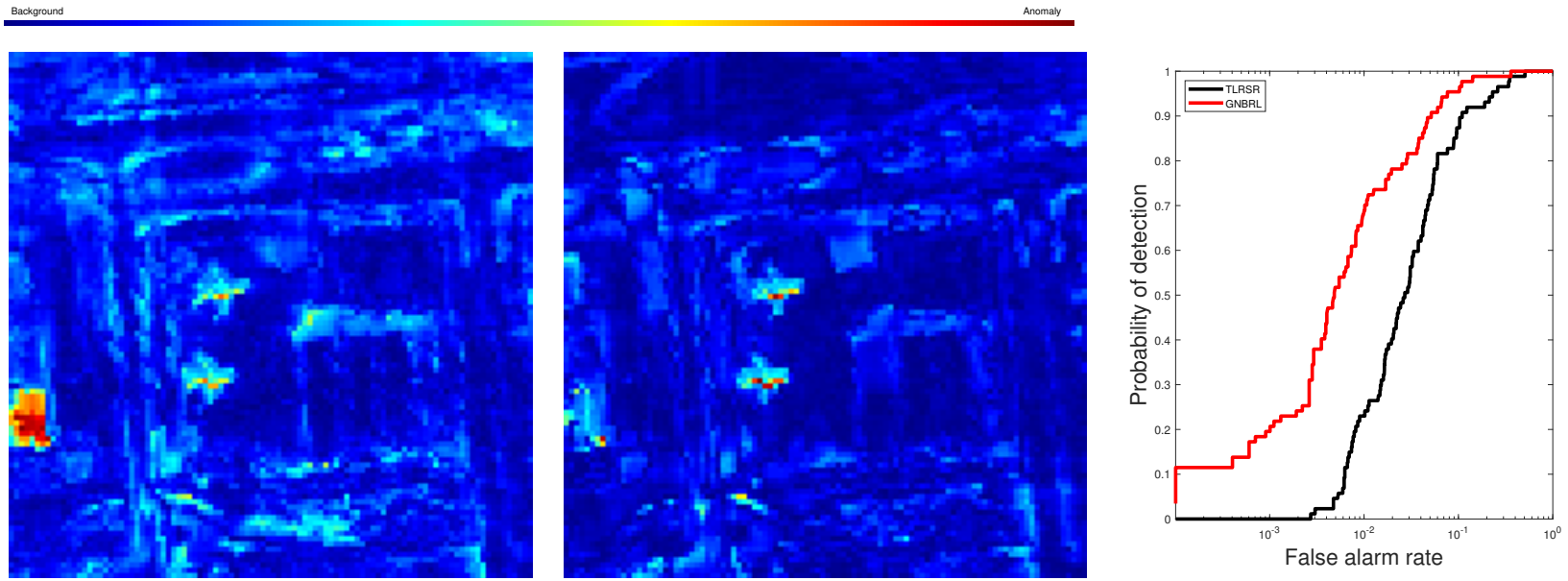
Performance of different nonconvex functions



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 3: AUC values (%) and corresponding running times of GNBRL with different nonconvex functions ψ for each data set.

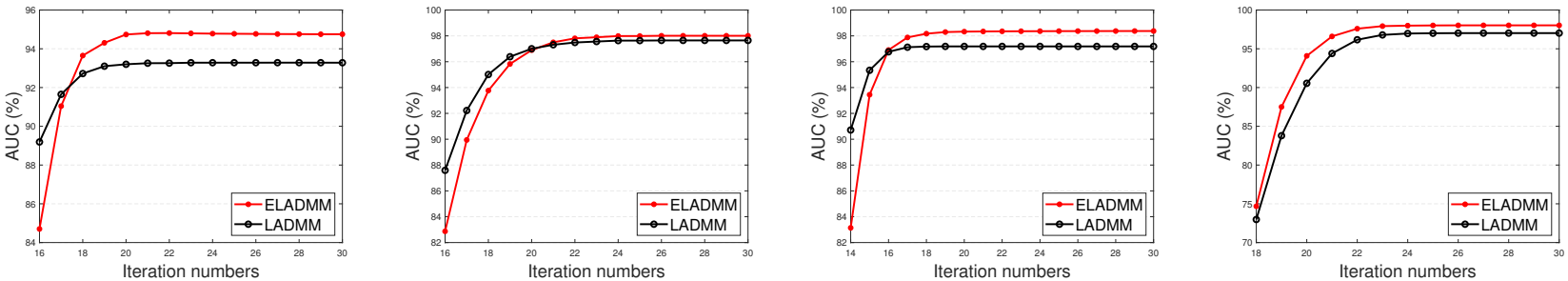
Effects of the dictionary constraint



(a) TLRSR (b) GNBRL (c) ROC curves

Figure 4: Detection maps and ROC curves obtained by TLRSR and GNBRL for the "Airport2" data set, respectively.

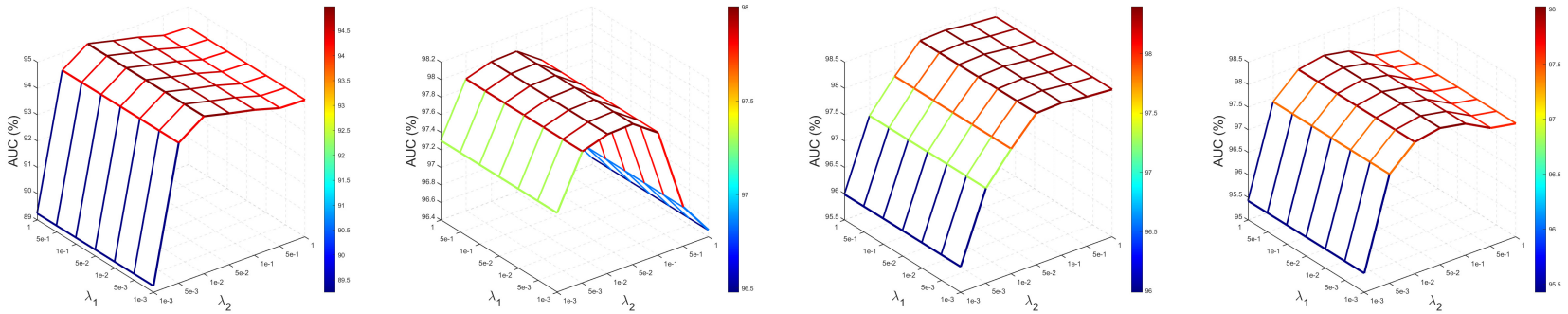
Effects of extrapolation strategy



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 5: AUC values (%) with respect to the iteration numbers for LADMM and ELADMM.

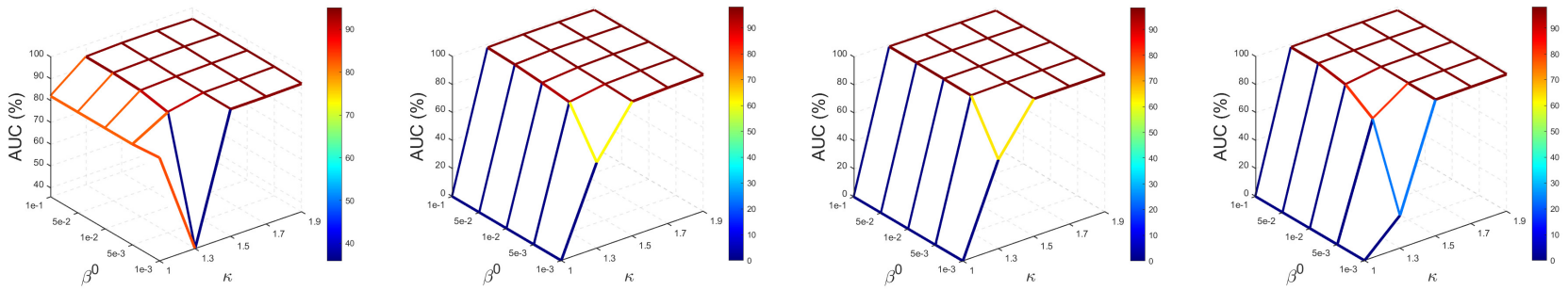
Parameters setting



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 6: Surfaces of AUC values (%) with different λ_1 and λ_2.

For simplicity, we set $\beta^{t+1} = \min\left\{\kappa\beta^t, 1e8\right\}$.



(a) Airport1 (b) Airport2 (c) Urban (d) Beach

Figure 7: Surfaces of AUC values (%) with different β^0 and κ.

Detection performance

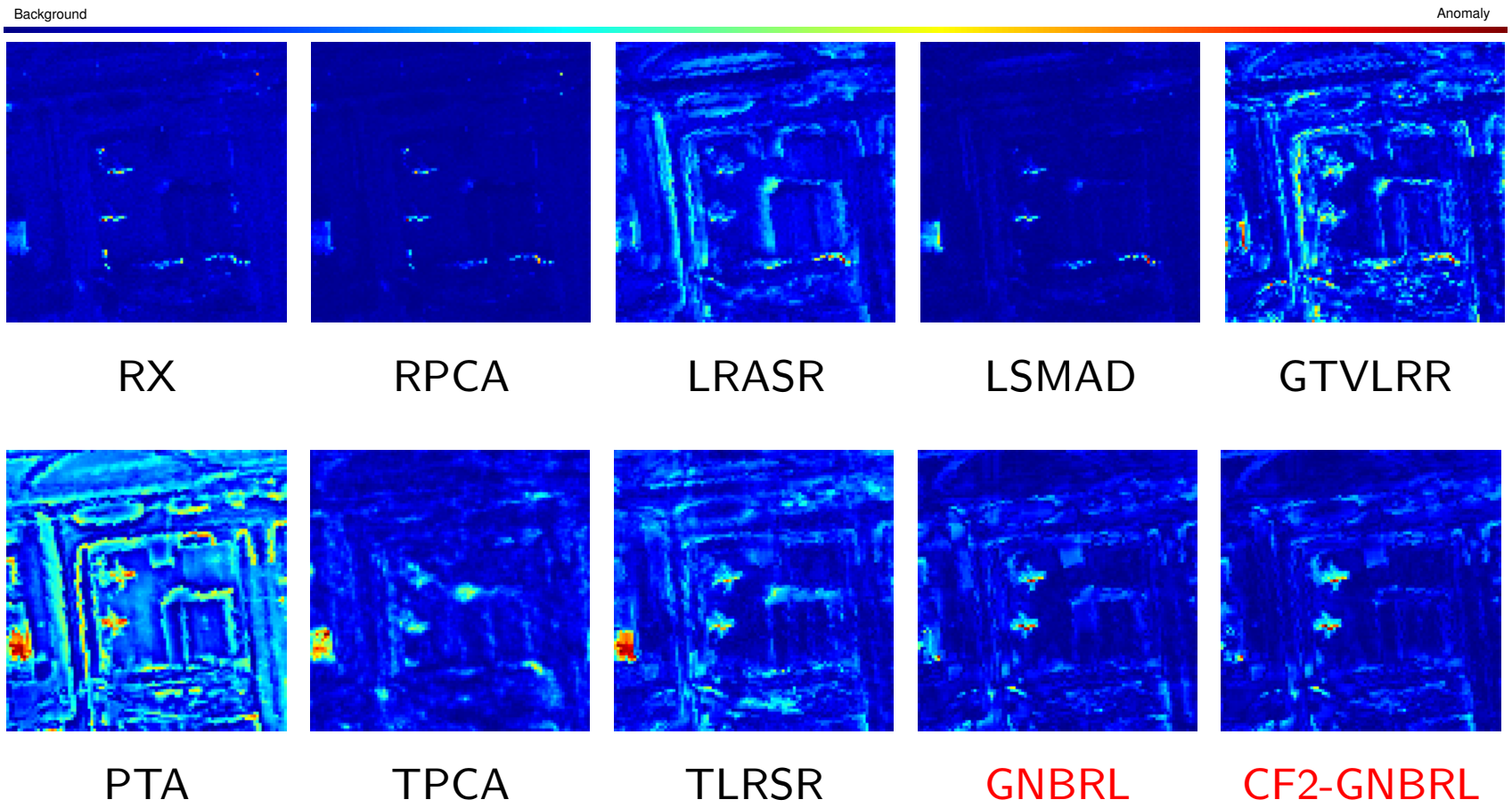
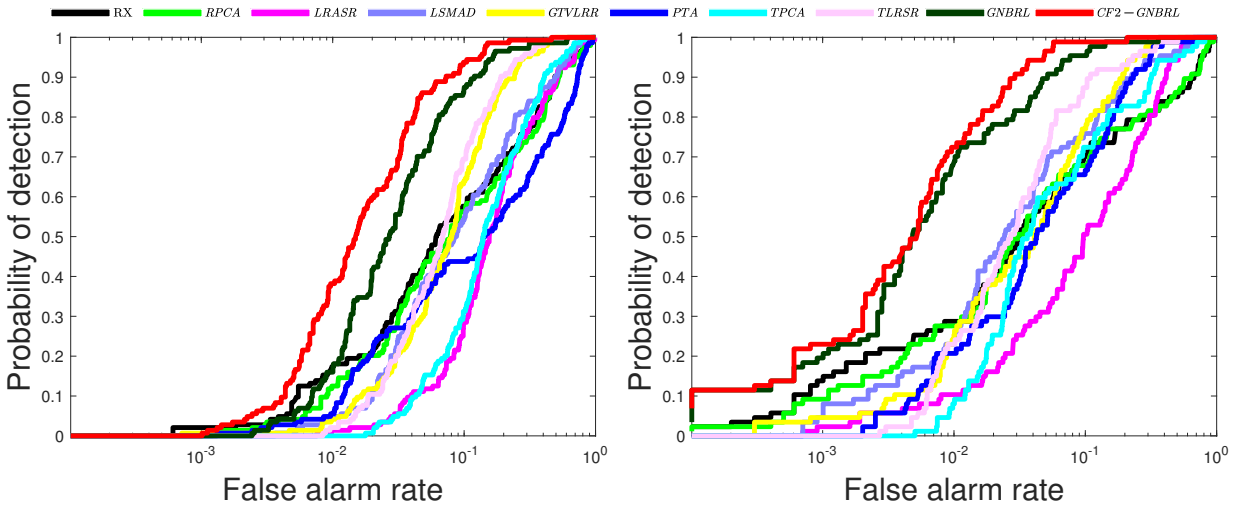
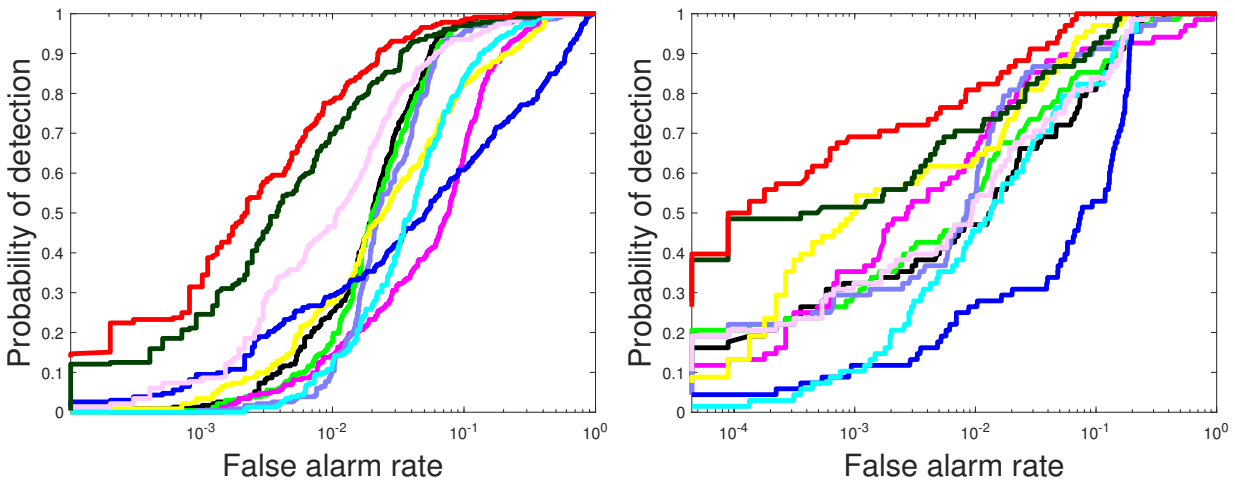


Figure 8: Target detection results by different methods.

ROC curves obtained by different methods for the four data sets.



(a) Airport1 (b) Airport2

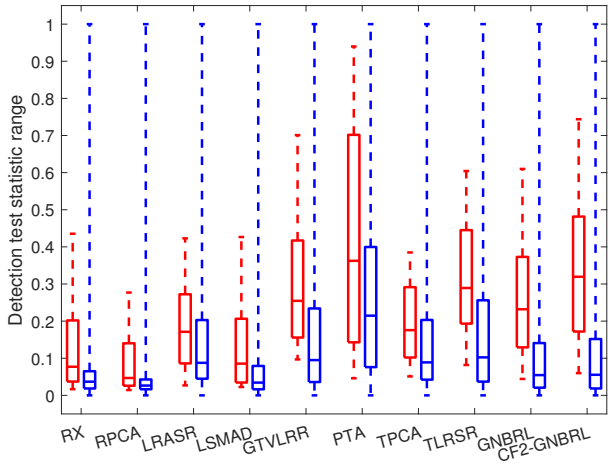


(c) Urban (d) Beach

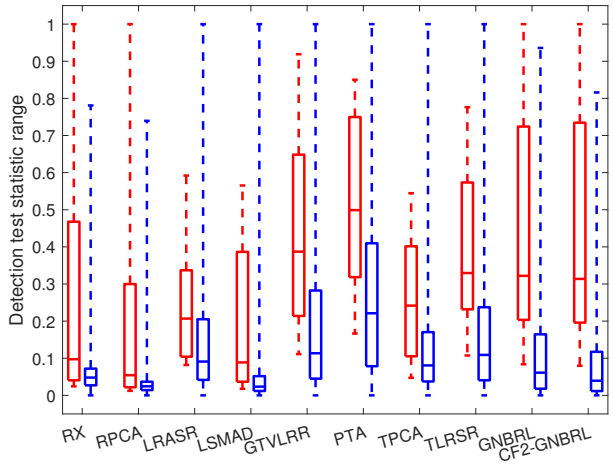
Table 1: Comparison of AUC values (%) and running time (s) of different methods for the four data sets.

HSI	Airport1		Airport2		Urban		Beach	
Algorithm	AUC	Time	AUC	Time	AUC	Time	AUC	Time
RX	82.21	0.42	84.03	0.41	96.92	0.41	95.39	0.04
RPCA	80.89	8.00	84.31	7.44	96.58	6.98	95.99	1.95
LRASR	77.28	53.81	86.48	70.13	92.89	47.51	95.65	104.90
LSMAD	83.39	9.54	92.17	8.60	96.05	8.74	97.06	7.65
GTVLRR	90.04	171.47	88.89	227.16	93.73	229.16	98.02	378.60
PTA	73.30	13.50	90.95	20.96	82.57	24.89	90.61	29.11
TPCA	80.22	30.91	88.90	30.62	93.69	22.15	95.82	21.71
TLRSR	90.56	3.44	94.57	3.63	97.10	3.58	95.98	5.84
GNBRL	94.75	1.60	98.00	1.50	98.38	1.91	98.03	4.01
CF2-GNBRL	**96.84**	27.14	**98.81**	31.63	**98.98**	31.40	**99.24**	83.06

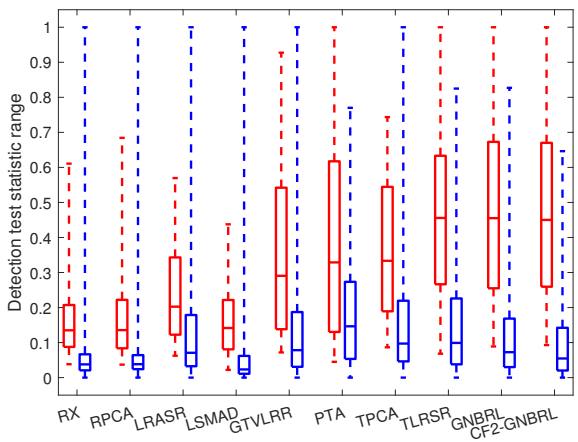
Separability maps of different methods for the four data sets.



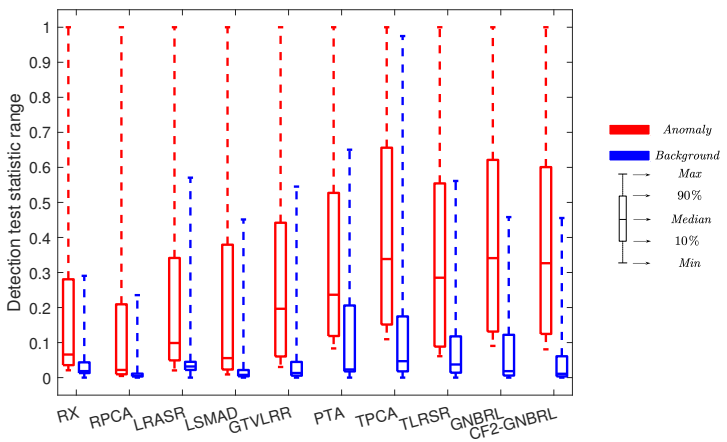
(a) Airport1



(b) Airport2



(c) Urban



(d) Beach

Contents

Our work

- We propose a GNBRL model that simultaneously learns the dictionary and anomaly tensor in a unified framework, which can enhance the quality of representation. By employing a class of generalized nonconvex functions as the TNN and l_1-norm approximations, we can capture the low rank structure of the background and the sparsity of the anomaly more accurately.
- We provide an error bounds of anomaly tensor recovered by the GNBRL model. Moreover, we develop an extrapolated linearized alternating direction method of multipliers (ELADMM) algorithm to solve the GNBRL model, and the convergence analysis is also given.

- We propose a coarse to fine two-stage (CF2) framework for GNBRL, which can greatly improve its performance by first obtaining the coarse anomaly tensor and then refining the anomaly tensor.

Thanks for your attention!