Volume Regularized Non-negative Matrix Factorizations

Andersen M. S. Ang and Nicolas Gillis,
Université de Mons, Belgium
{manshun.ang, nicolas.gillis}@umons.ac.be

Volume regularized NMF

Given $\mathbf{X} \in \mathbb{R}_+^{m\times n}$, find $\mathbf{W} \in \mathbb{R}_+^{m\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{n\times r}$ by solving

$$[\mathbf{W},\ \mathbf{H}] = \underset{\mathbf{W},\ \mathbf{H}}{\text{argmin}} \frac{1}{2}\|\mathbf{X} - \mathbf{W}\mathbf{H}^\top\| + \lambda g(\mathbf{W}) \text{ s.t. } \mathbf{W} \geq 0,\ \mathbf{H} \geq 0,\ \mathbf{H}\mathbf{1}_r = \mathbf{1}_r.$$

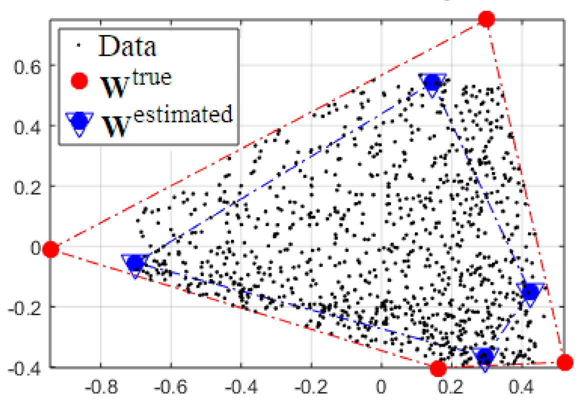
Here $g(\mathbf{W})$ are regularizers on the vol. of the cvx hull spanned by the columns of $\mathbf{W}$

$$g_{\det}(\mathbf{W}) = \frac{1}{2}\det(\mathbf{W}^\top\mathbf{W}), \quad g_{\text{logdet}}(\mathbf{W}) = \frac{1}{2}\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r).$$

We propose a new algorithm based on an eigenvalue upper bound of the logdet function.

Geometric interpretation :

$\mathbf{H}\mathbf{1}_r = \mathbf{1}_r$ means $\mathbf{H}$ row-stochastic, so the convex hull of the columns of $\mathbf{W}$ should approximate the columns of $\mathbf{X}$.
We assumes no pure data point.



Volume regularizers and algorithms

We use coorindate descent algorithms.
Update $\mathbf{H}$ by fast projected gradient method (FGM).
Update $\mathbf{W}$ based on the structure of g, using projected gradient descent method :

For the **Det regularizer** $\det(\mathbf{W}^\top\mathbf{W})$,
we optimize the column $\mathbf{w}_i$ of $\mathbf{W}$ sequentially [2]. We have

$$\det(\mathbf{W}^\top\mathbf{W}) = \gamma_i \mathbf{w}_i^\top \mathbf{B}_i \mathbf{w}_i,$$

where $\gamma_i = \det(\mathbf{W}_{\neq i}^\top \mathbf{W}_{\neq i})$, $\mathbf{B}_i = \mathbf{I}_m - \mathbf{W}_{\neq i}(\mathbf{W}_{\neq i}^\top\mathbf{W}_{\neq i})^{-1}\mathbf{W}_{\neq i}$.
Weighting $\mathbf{B}_i$ is proj. onto orthogonal complement of column space of $\mathbf{W}_{\neq i}$.
Interpretation : det = reweighted l_2 norm regularization on columns.

For the **logdet regularizer** $\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r)$,
we tackle using upper bounds :

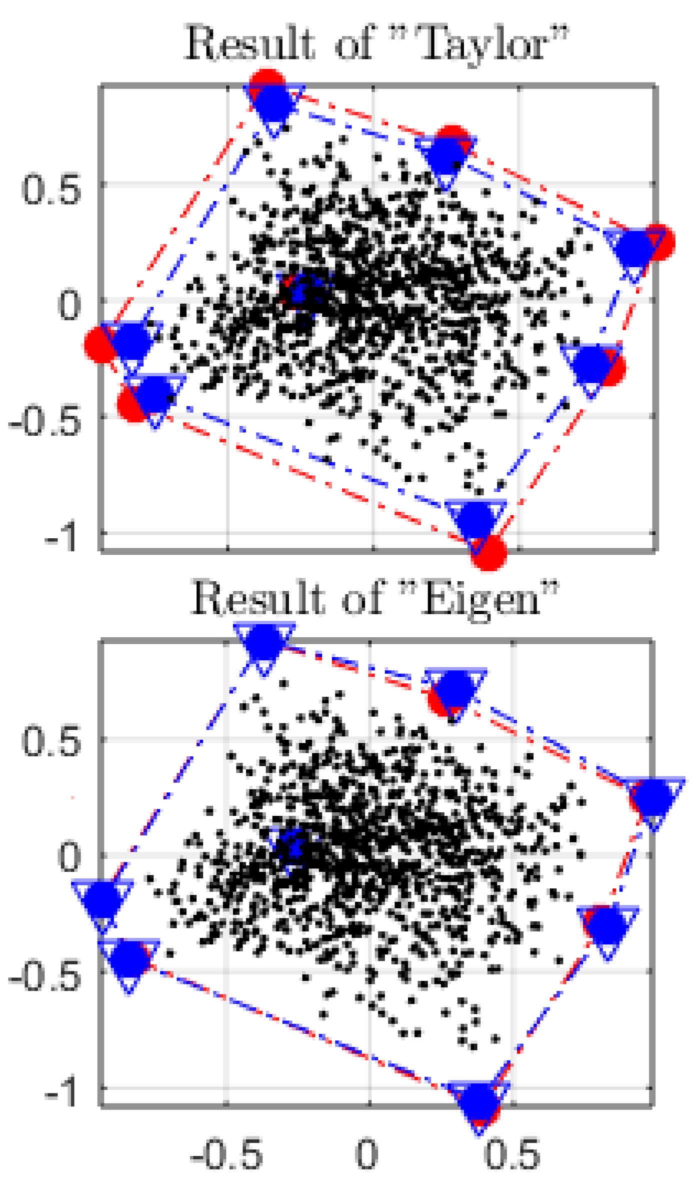
- **Taylor upper bound** $\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r) \leq \text{tr}(\mathbf{D}\mathbf{W}^\top\mathbf{W}) + c$, which is the first-order Taylor approximation around matrix $\mathbf{Y}$ where $\mathbf{D} = (\mathbf{Y}^\top\mathbf{Y} + \delta\mathbf{I}_r)^{-1}$.
  The weighting $\mathbf{D}$ is a dense matrix connecting columns of $\mathbf{W}$.
  It is a tight bound, with theoretical convergence guarantee.
  We use projected gradient for updating $\mathbf{W}$.
- **Eigen upper bound** $\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r) \leq \nu\text{tr}(\mathbf{W}^\top\mathbf{W}) + c = \nu\sum_i \|\mathbf{w}_i\|_2^2 + c$, where weighting $\nu = \left(\nu_{\min}(\mathbf{Y}^\top\mathbf{Y} + \delta\mathbf{I}_r)\right)^{-1}$ is a scalar.
  It is an eigen-approximation of Taylor bound.
  It is a loose bound, no theoretical convergence guarantee, but decomposable structure.
  We use column-wise update for updating $\mathbf{W}$: same as det.

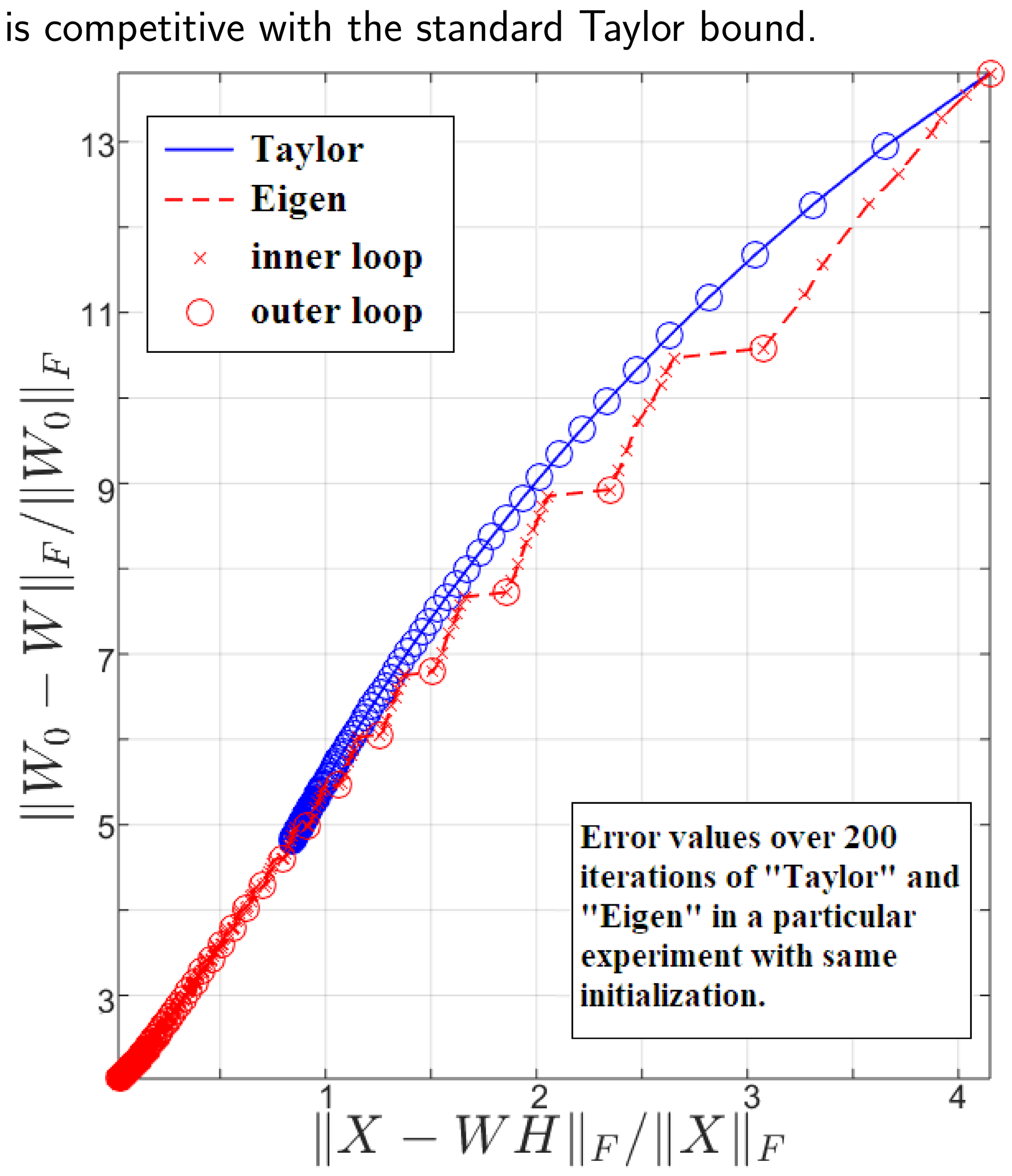
Experiments and results

Synthetic experiment - given (m, n, r) :

- Generate ground truth $\mathbf{W}_{ij} \sim \mathcal{U}[0\ 1]$ for all (i, j)
- Generate $\mathbf{H} \sim$ Dirichlete$(1, 1, 1)$ such that $\mathbf{H}$ is row-stochastic
- Observed data $\mathbf{X} = \mathbf{W}\mathbf{H} + \mathbf{N}$, $\mathbf{N}$ is white Gaussian noise
- Data points with corresponding $\mathbf{H}_{ij} > \theta$ are removed, $\theta \in [0\ 1]$ is purity index
- Initialize $\mathbf{W}^{\text{ini}}$, $\mathbf{H}^{\text{ini}}$ by SNPA, $\lambda = 5\dfrac{f(\mathbf{W}^{\text{ini}}, \mathbf{H}^{\text{ini}})}{|g(\mathbf{W}^{\text{ini}})|}$, 200 iterations

Result 1. The new algorithm "Eigen" is competitive with the standard Taylor bound.





Result 2. logdet regularizers work better than det regularizer.

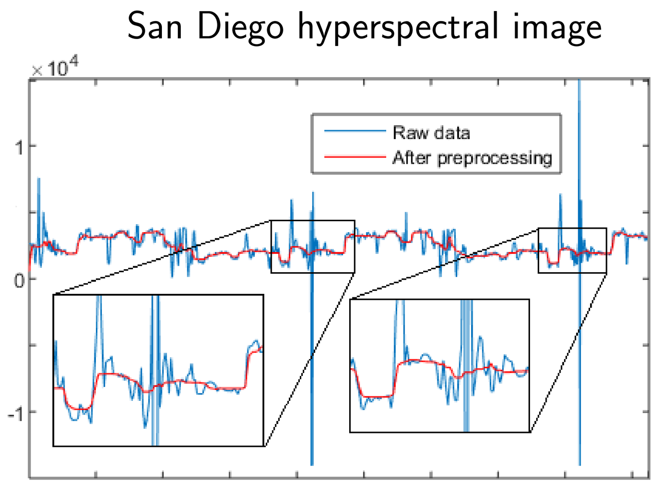
Table. Relative percentage error (mean±std) over 100 trials on fitting $\mathbf{X}$ (first column) and $\mathbf{W}$ (second column).

	$\theta = 0.9$, no noise			$\theta = 0.9$, 10% noise	
Det	2.49 ± 0.51	9.79 ± 1.49	Det	27.18 ± 0.45	36.64 ± 3.45
Taylor	0.46 ± 0.12	3.29 ± 0.64	Taylor	27.76 ± 0.33	$\mathbf{25.43 \pm 2.37}$
Eigen	$\mathbf{0.01 \pm 0.00}$	$\mathbf{1.19 \pm 0.40}$	Eigen	$\mathbf{23.64 \pm 0.14}$	33.21 ± 5.25
	$\theta = 0.7$, no noise			$\theta = 0.7$, 10% noise	
Det	3.36 ± 0.62	11.74 ± 2.05	Det	27.17 ± 0.42	39.03 ± 3.51
Taylor	1.76 ± 0.34	8.63 ± 1.13	Taylor	28.00 ± 0.34	$\mathbf{27.97 \pm 2.10}$
Eigen	$\mathbf{0.02 \pm 0.01}$	$\mathbf{2.80 \pm 1.50}$	Eigen	$\mathbf{23.58 \pm 0.14}$	37.43 ± 4.10

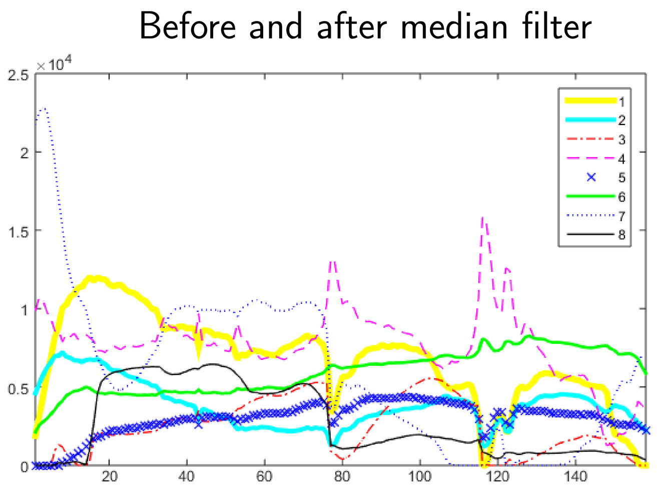
San Diego airport hyperspectral image

- "Eigen" with $r = 8$ (other parameters same as the synthetic cases).
- Preprocessing : replace negative values to 0, remove spikes by length-20 median filter.

San Diego hyperspectral image

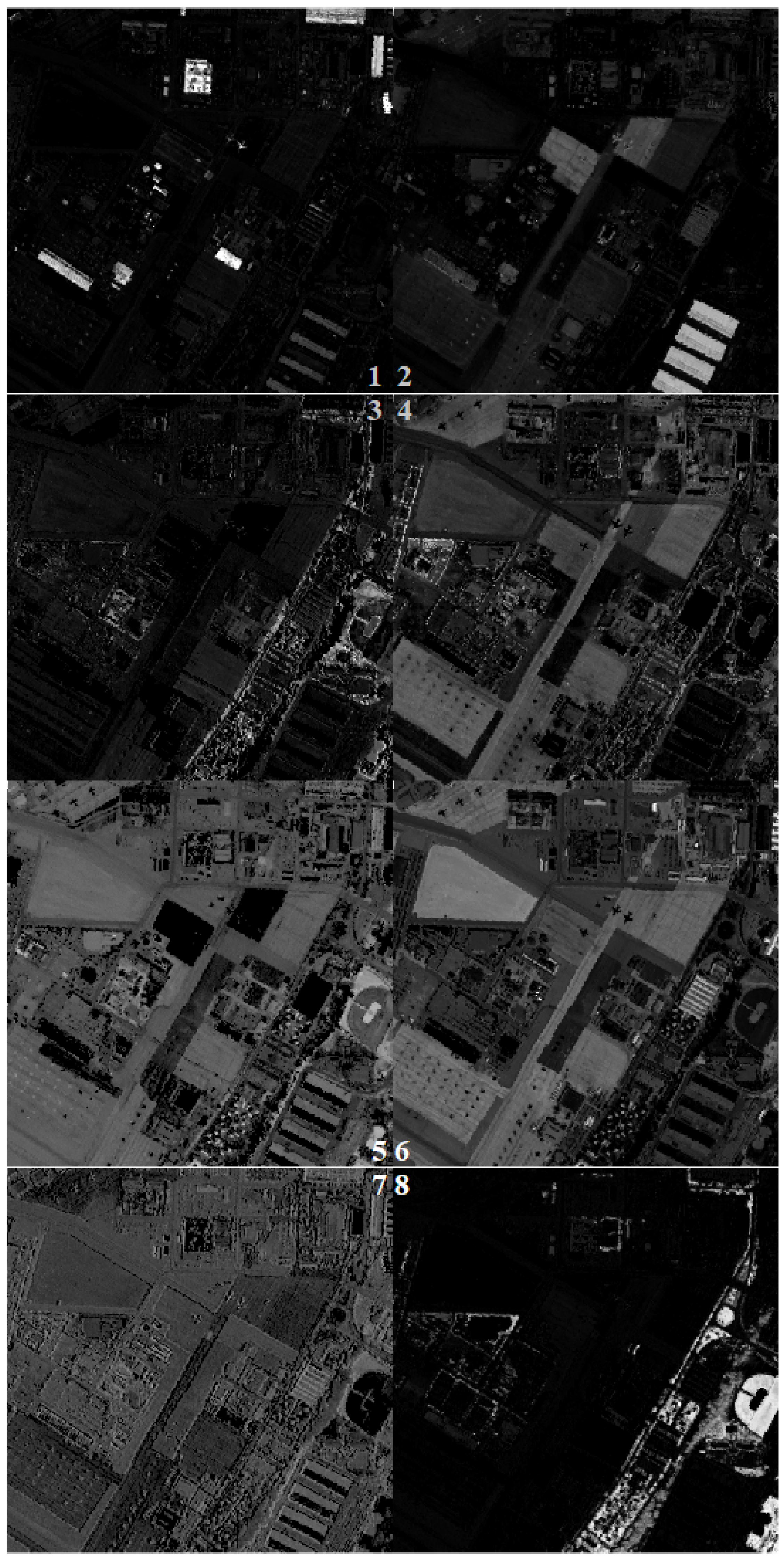


Before and after median filter



$\mathbf{W}$, the spectra extracted

- components 1, 2 : roof tops
- components 3, 8 : trees, grass
- others : different road surfaces



Conclusion

- logdet regularizers perform better than the det regularizer.
- a new algorithm using column-wise update of columns of $\mathbf{W}$ called "Eigen" for logdet regularizer.
- "Eigen" has a better numerical performance than the matrix-wise update algorithm "Taylor".
- "Eigen" can decompose the San Diego airport image into meaningful components.

References

[1] Andersen M.S. Ang and Nicolas Gillis, "Volume regularized Non-negative Matrix Factorizations", IEEE Whispers 2018, 23-26 September 2018, Amsterdam, Netherlands

[2] Zhou, G., Xie, S., Yang, Z., Yang, J.M. and He, Z., "Minimum-volume-constrained nonnegative matrix factorization: Enhanced ability of learning parts", IEEE Trans. NN, vol. 22, no. 10, pp.1626–1637, 2011

[3] Nicolas Gillis and Stephen Vavasis, "Fast and robust recursive algorithmsfor separable nonnegative matrix factorization", IEEE Trans. PAMI, vol. 36, no. 4, pp.698–714, 2014