rSVDdpd

A Robust Singular Value Decomposition Technique

Ayanendranath Basu

Interdisciplinary Statistical Research Unit (ISRU),
Indian Statistical Institute, Kolkata, India

EcoSta - 2024

Background Modelling Problem

Background Modelling Problem

=

+

Background Modelling Problem

=

+

Applications ranging security, defence, object tracking, motion detection, video filters, etc.

Background Modelling as Low Rank Decomposition

$h \times w \times t$

Background Modelling as Low Rank Decomposition

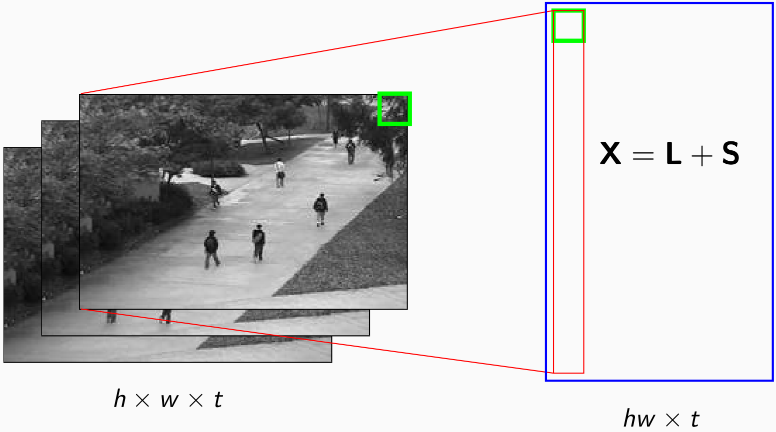
$h \times w \times t$

Background Modelling as Low Rank Decomposition

$h \times w \times t$

$hw \times t$

Background Modelling as Low Rank Decomposition



Singular Value Decomposition (SVD)

Definition
For a $n \times p$ matrix $\mathbf{M}$, its singular value decomposition is defined by the factorization

$$\mathbf{M} = \mathbf{U}_{n \times r} \mathbf{D}_{r \times r} (\mathbf{V}_{p \times r})^{\intercal}$$

where,

1. $\mathbf{U}, \mathbf{V}$ are unitary matrices (left and right singular vectors).
2. $\mathbf{D}$ is a diagonal matrix comprising singular values, in decreasing order of magnitude.
3. r is the approximate rank.

Robust Background Modelling

- Singular Value Decomposition (SVD) can be used to estimate **L**.
- Advantages compared to Deep learning.
    1. Unsupervised. Need no training data.
    2. Low hardware requirements.
    3. Generalizable.
    4. Better theoretical guarantees.

Robust Background Modelling

- Singular Value Decomposition (SVD) can be used to estimate **L**.
- Advantages compared to Deep learning.
  1. Unsupervised. Need no training data.
  2. Low hardware requirements.
  3. Generalizable.
  4. Better theoretical guarantees.

- Real-life surveillance has many nuisances.
  1. Weather conditions (Fog / Rain)

Robust Background Modelling

- Singular Value Decomposition (SVD) can be used to estimate **L**.
- Advantages compared to Deep learning.
  1. Unsupervised. Need no training data.
  2. Low hardware requirements.
  3. Generalizable.
  4. Better theoretical guarantees.

- Real-life surveillance has many nuisances.
  1. Weather conditions (Fog / Rain)
  2. Naturally moving background

Robust Background Modelling

- Singular Value Decomposition (SVD) can be used to estimate **L**.
- Advantages compared to Deep learning.
  1. Unsupervised. Need no training data.
  2. Low hardware requirements.
  3. Generalizable.
  4. Better theoretical guarantees.

- Real-life surveillance has many nuisances.
  1. Weather conditions (Fog / Rain)
  2. Naturally moving background
  3. Camera tampering

SVD as alternating regression

$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$; $\mathbf{L}$ low-rank, $\mathbf{S}$ sparse and $\mathbf{N}$ dense noise

SVD as alternating regression

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N};\ \ \mathbf{L} \text{ low-rank}, \mathbf{S} \text{ sparse and } \mathbf{N} \text{ dense noise}$$

$$\Rightarrow \quad \mathbf{X} = \mathbf{L} + \mathbf{E},\ \text{where, } \mathbf{E} = ((\epsilon_{ij}))_{i,j};\ \ \epsilon_{ij} \sim (1-\delta)g + \delta h$$

SVD as alternating regression

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N};\ \ \mathbf{L} \text{ low-rank}, \mathbf{S} \text{ sparse and } \mathbf{N} \text{ dense noise}$$

$$\Rightarrow \quad \mathbf{X} = \mathbf{L} + \mathbf{E},\ \text{where, } \mathbf{E} = ((\epsilon_{ij}))_{i,j};\ \epsilon_{ij} \sim (1-\delta)g + \delta h$$

$$\Rightarrow \quad X_{ij} = \sum_{k=1}^{r} \lambda_k u_{ki} v_{kj} + \epsilon_{ij}$$

SVD as alternating regression

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}; \;\; \mathbf{L} \text{ low-rank}, \mathbf{S} \text{ sparse and } \mathbf{N} \text{ dense noise}$$

$$\Rightarrow \quad \mathbf{X} = \mathbf{L} + \mathbf{E}, \;\text{ where, } \mathbf{E} = ((\epsilon_{ij}))_{i,j}; \;\; \epsilon_{ij} \sim (1-\delta)g + \delta h$$

$$\Rightarrow \quad X_{ij} = \sum_{k=1}^{r} \lambda_k u_{ki} v_{kj} + \epsilon_{ij}$$

$$\Rightarrow \quad X_{ij} = \sum_{k=1}^{r} a_{ki} b_{kj} + \epsilon_{ij}; \;\text{ assume, } a_{ki} = \sqrt{\lambda_k} u_{ki}, b_{kj} = \sqrt{\lambda_k} v_{kj}$$

SVD as alternating regression

$$
\begin{aligned}
& \mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N};\ \ \mathbf{L} \text{ low-rank}, \mathbf{S} \text{ sparse and } \mathbf{N} \text{ dense noise} \\
\Rightarrow\quad & \mathbf{X} = \mathbf{L} + \mathbf{E},\ \text{where, } \mathbf{E} = ((\epsilon_{ij}))_{i,j};\ \epsilon_{ij} \sim (1-\delta)g + \delta h \\
\Rightarrow\quad & X_{ij} = \sum_{k=1}^{r} \lambda_k u_{ki} v_{kj} + \epsilon_{ij} \\
\Rightarrow\quad & X_{ij} = \sum_{k=1}^{r} a_{ki} b_{kj} + \epsilon_{ij};\ \text{assume, } a_{ki} = \sqrt{\lambda_k} u_{ki}, b_{kj} = \sqrt{\lambda_k} v_{kj}
\end{aligned}
$$

Fixing j yields,

$$
\begin{bmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{nj} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{r1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{rn} \end{bmatrix}_{n \times r} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{rj} \end{bmatrix} + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \vdots \\ \epsilon_{nj} \end{bmatrix}
$$

Density Power Divergence

Definition
Given two density functions f and g, Basu et al. (1998) defines the density power divergence between them as

$$d_\alpha(g, f) = \begin{cases} \int f^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \int f^\alpha g + \frac{1}{\alpha} \int g^{1+\alpha}, & \alpha > 0 \\ \int f \log(f/g) & \alpha = 0 \end{cases}$$

Density Power Divergence

Definition
Given two density functions f and g, Basu et al. (1998) defines the density power divergence between them as

$$d_\alpha(g, f) = \begin{cases} \int f^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \int f^\alpha g + \frac{1}{\alpha} \int g^{1+\alpha}, & \alpha > 0 \\ \int f \log(f/g) & \alpha = 0 \end{cases}$$

Definition
Given an independent sample of observations $X_1, \ldots X_n$ modelled by non-homogeneous family of distributions $\{f_{i,\theta} : \theta \in \Theta\}$, the MDPDE (Ghosh and Basu) is defined as

$$\widehat{\theta}_\alpha = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \left[\int f_{i,\theta}^{1+\alpha} - \left(1 + \alpha^{-1}\right) f_{i,\theta}^{\alpha}(X_i) \right]$$

Robust SVD using DPD

For background modelling problem, usually $r = 1$, so we have

$$X_{ij} = a_i b_j + \epsilon_{ij}$$

So we choose $f_{ij,\theta} \equiv N(a_i b_j, \sigma^2)$.

Robust SVD using DPD

For background modelling problem, usually $r = 1$, so we have

$$X_{ij} = a_i b_j + \epsilon_{ij}$$

So we choose $f_{ij,\theta} \equiv N(a_i b_j, \sigma^2)$.

$$w_{ij}^{(t)} = \exp\left\{-\alpha(X_{ij} - a_i^{(t)} b_j^{(t)})^2/(\sigma^{(t)})^2\right\}$$

Robust SVD using DPD

For background modelling problem, usually $r = 1$, so we have

$$X_{ij} = a_i b_j + \epsilon_{ij}$$

So we choose $f_{ij,\theta} \equiv N(a_i b_j, \sigma^2)$.

$$\begin{aligned} w_{ij}^{(t)} &= \exp\left\{ -\alpha (X_{ij} - a_i^{(t)} b_j^{(t)})^2 / (\sigma^{(t)})^2 \right\} \\ a_i^{(t+1/2)} &= \sum_j (x_{ij}/b_j^{(t)})(b_j^{(t)})^2 w_{ij}^{(t)} / \sum_j (b_j^{(t)})^2 w_{ij}^{(t)}, \ i = 1, \dots n; \end{aligned}$$

Robust SVD using DPD

For background modelling problem, usually $r = 1$, so we have

$$X_{ij} = a_i b_j + \epsilon_{ij}$$

So we choose $f_{ij,\theta} \equiv N(a_i b_j, \sigma^2)$.

$$\begin{aligned}
w_{ij}^{(t)} &= \exp\left\{-\alpha(X_{ij} - a_i^{(t)} b_j^{(t)})^2/(\sigma^{(t)})^2\right\} \\
a_i^{(t+1/2)} &= \sum_j (x_{ij}/b_j^{(t)})(b_j^{(t)})^2 w_{ij}^{(t)} / \sum_j (b_j^{(t)})^2 w_{ij}^{(t)},\ \ i = 1, \ldots n; \\
a_i^{(t+1)} &= a_i^{(t+1/2)}/|a_i^{(t+1/2)}|
\end{aligned}$$

Robust SVD using DPD

For background modelling problem, usually $r = 1$, so we have

$$X_{ij} = a_i b_j + \epsilon_{ij}$$

So we choose $f_{ij,\theta} \equiv N(a_i b_j, \sigma^2)$.

$$\begin{aligned}
w_{ij}^{(t)} &= \exp\left\{-\alpha(X_{ij} - a_i^{(t)} b_j^{(t)})^2/(\sigma^{(t)})^2\right\} \\
a_i^{(t+1/2)} &= \sum_j (x_{ij}/b_j^{(t)})(b_j^{(t)})^2 w_{ij}^{(t)} / \sum_j (b_j^{(t)})^2 w_{ij}^{(t)},\ i = 1, \dots n; \\
a_i^{(t+1)} &= a_i^{(t+1/2)}/|a_i^{(t+1/2)}| \\
b_j^{(t+1)} &= \sum_i (x_{ij}/a_i^{(t+1)})(a_i^{(t+1)})^2 w_{ij}^{(t)} / \sum_i (a_i^{(t+1)})^2 w_{ij}^{(t)},\ j = 1, \dots p; \\
(\sigma^{(t+1)})^2 &= \sum_i \sum_j (x_{ij} - a_i^{(t+1)} b_j^{(t+1)})^2 w_{ij}^{(t)} / \left(\sum_i \sum_j w_{ij}^{(t)} - \frac{\alpha}{(1+\alpha)^{3/2}}\right).
\end{aligned}$$

Features and Theoretical Results (Roy, Basu and Ghosh, 2021)

1. Massively parallelizable iteration rules.
2. Converges to a global optimum of the restricted DPD objective function.
3. Estimator has equivariance property under general g.
4. Asymptotically consistent under reasonable assumptions:
    4.1 Thrice differentiability of density g and exchange of derivative under integral sign.
    4.2 Lindeberg-like condition on third derivative of density function g.
    4.3 Both $n = hw, p = t \to \infty$ with $n/p \to c$ for some $0 < c < \infty$.
    4.4 $\sigma = O((np)^{-1/2})$, i.e., $\sigma^2 \to 0$ as $n, p \to \infty$.
5. When $p = t \to \infty$, asymptotic breakdown point is $\alpha/(1+\alpha)$, free of $n = hw$. Very robust and attractive for large resolution images (Roy, Ghosh and Basu, 2023).

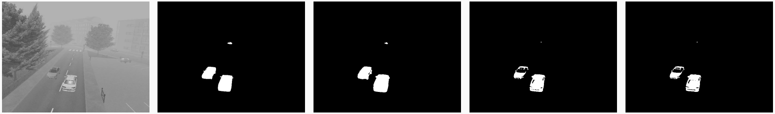
Benchmarks

- We consider competing algorithms like GoDec, GRASTA, Variational Bayes, Outlier Pursuit, PCP, Sparse PCP, ADMM, ALM, etc.
- rSVDdpd is about 50x faster than PCP, and 5-10x faster than all other algorithms. Following are comparisons on some videos in **Background Model Challenger** Dataset.

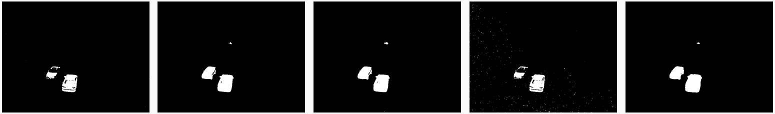
Dataset	Metric	rSVDdpd	GoDec	GRASTA	VB	PCP
BMC (street)	F1	0.754	0.672	0.528	0.728	0.739
BMC (rotary)	F1	0.757	0.723	0.624	0.614	0.726

- rSVDdpd has applications beyonds background modelling, can be used as a general purpose robust SVD.

Background Model Challenger Dataset



(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)

One frame of the street video (a), ground truth (b) and estimated foreground mask from exact RPCA (c), inexact ALM (d), SRPCP (e), Variational Bayes (f), Outlier Pursuit (g), GoDec (h), GRASTA (i) and rSVDdpd (j) algorithms. Enlarged better quality images are available online at https://subroy13.github.io/rsvddpd-home/.

University of Houston Camera Tampering Dataset

Truth OP ADMM GoDec rSVDdpd

References

- Basu, Ayanendranath, et al. "Robust and efficient estimation by minimising a density power divergence." Biometrika 85.3 (1998): 549-559.
- Ghosh, Abhik, and Ayanendranath Basu. "Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression." (2013): 2420-2456.
- Roy, Subhrajyoty, Ayanendranath Basu, and Abhik Ghosh. "A New Robust Scalable Singular Value Decomposition Algorithm for Video Surveillance Background Modelling." arXiv preprint arXiv:2109.10680 (2021).
- Roy, Subhrajyoty, Abhik Ghosh, and Ayanendranath Basu. "Analysis of the rSVDdpd Algorithm: A Robust Singular Value Decomposition Method using Density Power Divergence." arXiv preprint arXiv:2307.10591 (2023).
- Roy, Subhrajyoty, et al. "Breakdown Point Analysis of the Minimum S-Divergence Estimator." arXiv preprint arXiv:2304.07466 (2023).
- R package `rsvddpd` `https://cran.r-project.org/web/packages/rsvddpd`.

Questions?