Supplemental material to: “Dictionary learning for signals in additive noise with generalized Gaussian distribution ”

Xiaomeng Zheng[a], Bogdan Dumitrescu[b], Jiamou Liu[c], and Ciprian Doru Giurcăneanu[a]

[a]Department of Statistics, Univ. of Auckland, New Zealand
[b]University Politehnica of Bucharest, Romania
[c]School of Computer Science, Univ. of Auckland, New Zealand

Contents

1 BIC formula for the linear model in additive Gaussian noise

We assume the following model for the measurements $\boldsymbol{y}$:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{u}, \tag{1}$$

where the entries of the matrix $\boldsymbol{H} \in \mathbb{R}^{m\times n}$ are *known*. Note that m stands for the number of measurements and n is the number of linear parameters. In our calculations, we suppose that $m > n$. The random vector $\boldsymbol{u}$ is assumed to be Gaussian distributed with zero-mean and covariance matrix $\tau\boldsymbol{I}$, where $\tau > 0$ and $\boldsymbol{I}$ denotes the identity matrix of appropriate dimension.

BIC is derived by considering an asymptotic approximation for the penalty term of the following criterion [7]:

$$-\mathcal{L}(\boldsymbol{y}\ ;\ \widehat{\boldsymbol{x}}, \widehat{\tau}) + \frac{1}{2}\ln|\widehat{\boldsymbol{J}}|, \tag{2}$$

where $\mathcal{L}(\cdot\ ;\ \cdot)$ denotes the log-likelihood function, $\ln(\cdot)$ is the notation for the natural logarithm, the operator $|\cdot|$ is used for the determinant of the matrix in the argument and $\widehat{\boldsymbol{J}}$ stands for the sample Fisher Information Matrix (FIM) in which the unknown values of the parameters are replaced with their maximum likelihood (ML) estimates. In our case, the ML estimates are given by the formulas $\widehat{\boldsymbol{x}} = (\boldsymbol{H}^\top\boldsymbol{H})^{-1}\boldsymbol{H}^\top\boldsymbol{y}$, $\widehat{\tau} = \dfrac{||\boldsymbol{y}-\boldsymbol{H}\widehat{\boldsymbol{x}}||_2^2}{m}$ [8]. Hence, we have:

$$\mathcal{L}(\boldsymbol{y}\ ;\ \widehat{\boldsymbol{x}}, \widehat{\tau}) = -\frac{m}{2}\ln\widehat{\tau} - \frac{m}{2}\ln(2\pi) - \frac{m}{2}. \tag{3}$$

The sample FIM has the expression: $\widehat{\boldsymbol{J}} = \begin{bmatrix} (\boldsymbol{H}^\top\boldsymbol{H})/\widehat{\tau} & \boldsymbol{0} \\ \boldsymbol{0} & m/(2\widehat{\tau}^2) \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{J}}_n & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{J}_\tau \end{bmatrix}$, where the size of the block $\widehat{\boldsymbol{J}}_n$ is $n\times n$ [8].

After discarding the terms that do not depend on n, the criterion in (2) can be re-written as $\dfrac{m}{2}\ln\widehat{\tau} + \dfrac{1}{2}\ln|\widehat{\boldsymbol{J}}|$. The fact that $\dfrac{1}{m}\widehat{\boldsymbol{J}} = \mathcal{O}(1)$ (when $m \gg 1$) leads to $\ln|\widehat{\boldsymbol{J}}| = \ln\left|m\dfrac{1}{m}\widehat{\boldsymbol{J}}\right| = (n+1)\ln m + \mathcal{O}(1)$. Hence, we get the well-known criterion, $\dfrac{m}{2}\ln\widehat{\tau} + \dfrac{n+1}{2}\ln m$, which in most of the cases is written as

$$\mathrm{BIC}_{\mathrm{Gauss}} \quad = \quad \frac{m}{2}\ln\widehat{\tau} + \frac{n}{2}\ln m. \tag{4}$$

2 Proof of Eq. (19) from [1]

For ease of writing, we define: $a = \dfrac{c_1(\beta)}{\sqrt{\tau}}$ and $b = \dfrac{c_2(\beta)}{\tau^{1/(1+\beta)}}$. For $u > 0$, the expression in [1, Eq. (16)] becomes $f(u\ ;\ \beta,\tau) = a\exp\left(-bu^{2/(1+\beta)}\right)$, which leads to

$$\left(\frac{\partial}{\partial u}f(u\ ;\ \beta,\tau)\right)^2 = a^2b^2\exp\left(-2bu^{2/(1+\beta)}\right)\left(\frac{2}{1+\beta}\right)^2 u^{2(1-\beta)/(1+\beta)}.$$

Furthermore, we have:

$$\frac{\left(\frac{\partial}{\partial u}f(u\ ;\ \beta,\tau)\right)^2}{f(u\ ;\ \beta,\tau)} = ab^2\exp\left(-bu^{2/(1+\beta)}\right)\left(\frac{2}{1+\beta}\right)^2 u^{2(1-\beta)/(1+\beta)}.$$

Therefore, we get

$$\begin{aligned}\mathbb{E}\left[\left(\frac{\partial}{\partial u}\ln f(u\ ;\ \beta,\tau)\right)^2\right] &= 2\int_0^\infty \frac{\left(\frac{\partial}{\partial u}f(u\ ;\ \beta,\tau)\right)^2}{f(u\ ;\ \beta,\tau)}\mathrm{d}u \\ &= 2ab^2\left(\frac{2}{1+\beta}\right)^2\int_0^\infty u^{2(1-\beta)/(1+\beta)}\exp\left(-bu^{2/(1+\beta)}\right)\mathrm{d}u.\end{aligned}$$

With the change of variable $z = u^{2/(1+\beta)}$, the integral above becomes:

$$\frac{1+\beta}{2}\int_0^\infty z^{(1-\beta)/2}\exp(-bz)\mathrm{d}z = \frac{1+\beta}{2}\frac{\Gamma\left(\frac{3-\beta}{2}\right)}{b^{(3-\beta)/2}}.$$

For evaluating the integral, we have applied [9, Theorem 5.7.3], with the supplementary assumption that $\beta < 3$. It follows that

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{\partial}{\partial u}\ln f(u\,;\,\beta,\tau)\right)^2\right] \\
&= ab^{(1+\beta)/2}\frac{4}{1+\beta}\Gamma\left(\frac{3-\beta}{2}\right) \\
&= \frac{\Gamma^{1/2}\left(\frac{3+3\beta}{2}\right)}{(1+\beta)\Gamma^{3/2}\left(\frac{1+\beta}{2}\right)\tau^{1/2}} \\
&\quad\cdot\frac{\Gamma^{1/2}\left(\frac{3+3\beta}{2}\right)}{\Gamma^{1/2}\left(\frac{1+\beta}{2}\right)\tau^{1/2}} \\
&\quad\cdot\frac{4}{1+\beta}\Gamma\left(\frac{3-\beta}{2}\right) \\
&= \frac{1}{\tau}\frac{4}{(1+\beta)^2}\frac{\Gamma\left(\frac{3+3\beta}{2}\right)\Gamma\left(\frac{3-\beta}{2}\right)}{\Gamma^2\left(\frac{1+\beta}{2}\right)}.
\end{aligned}$$

This result together with [1, Eq. (18)] lead to [1, Eq. (19)].

3 Graphical illustration for the penalty term in the formula of BIC_β in [1, Eq. (20)]

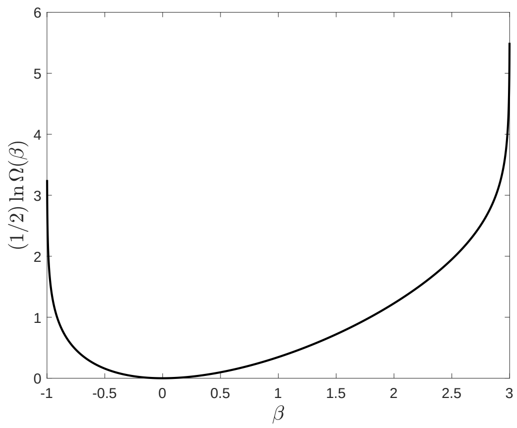

Figure 1: The extra-cost for each linear parameter, in comparison with the Gaussian case, when the shape parameter β takes values in the interval $(-1, 3)$.

4 Experiments with simulated data

4.1 Results for the case when the sparsity level is known

We show in Tables 1-2 the statistics for $\widehat{n}$ (minimum, average and maximum) collected from ten runs, for both learning algorithms ADL and ADL_1. Based on the results reported in the tables, we have: (i) In the case when the additive noise has Gaussian distribution, ADL_1/EBIC works better than all other combinations in selecting the dictionary sizes that are closer to the true dictionary size for all noise levels. However, when the additive noise is Laplacian distributed, ADL_1/BIC is slightly better than the other methods; (ii) In most of the cases, the maximum values of $\widehat{n}$ chosen by the two IT criteria when they are used together with ADL_1 algorithm are much smaller than the maximum values of $\widehat{n}$ that are selected when the ADL algorithm is used.

Gaussian Noise		ADL/EBIC	ADL_1/EBIC	ADL/BIC	ADL_1/BIC
SNR=20 dB	min	50.0	50.0	50.0	50.0
	avg	51.7	51.6	52.9	**50.6**
	max	56.0	55.0	63.0	55.0
SNR=40 dB	min	50.0	50.0	50.0	50.0
	avg	69.0	**50.3**	79.4	51.5
	max	123.0	51.0	134.0	58.0
SNR=60 dB	min	50.0	50.0	50.0	50.0
	avg	66.8	**50.4**	85.4	52.6
	max	105.0	51.0	152.0	58.0

Table 1: Statistics for the dictionary sizes selected in 10 runs. In each case, we report the minimum, the average and the maximum values of $\widehat{n}$. For each SNR, we write in bold the average value of $\widehat{n}$ that is closest to $n_{true} = 50$. In this experiment, the sparsity level is assumed to be known: $s = s_{true} = 3$.

Laplacian Noise		ADL/EBIC	ADL_1/EBIC	ADL/BIC	ADL_1/BIC
SNR=20 dB	min	50.0	50.0	50.0	50.0
	avg	51.9	52.1	53.3	**50.7**
	max	61.0	61.0	59.0	55.0
SNR=40 dB	min	50.0	50.0	51.0	50.0
	avg	62.3	53.0	72.3	**52.0**
	max	104.0	56.0	144.0	61.0
SNR=60 dB	min	50.0	50.0	51.0	50.0
	avg	59.1	**50.8**	81.5	51.9
	max	73.0	57.0	125.0	60.0

Table 2: All the settings are as in Table 1, except that the distribution of the additive noise is Laplacian.

4.2 Results for the case when the sparsity level is unknown

For evaluating the performance of BIC and EBIC in selecting both the sparsity level $\widehat{s}$ and the dictionary size $\widehat{n}$, we show in Tables 3-4 the minimum, average and maximum values of $\widehat{s}$ and $\widehat{n}$ collected from ten runs. It is easy to observe the following: (i) ADL_1/EBIC works better than all other combinations in selecting the sparsity level $\widehat{s}$, which is equal to the true sparsity level ($s_{true} = 3$) for all types of noise and noise levels; (ii) When the additive noise has Gaussian distribution, ADL_1/BIC is superior to ADL_1/EBIC in selecting the dictionary size at the noise level that equals 40dB. By contrast, ADL_1/EBIC is slightly better when the noise level is 60dB; (iii) When the additive noise is Laplacian distributed, ADL_1/BIC works slightly better than ADL_1/EBIC in selecting the dictionary size, in the cases of noise level greater than 20dB.

Gaussian Noise		ADL/EBIC	ADL_1/EBIC	ADL/BIC	ADL_1/BIC
SNR=20 dB	min	3.0 50.0	3.0 50.0	3.0 50.0	3.0 50.0
	avg	**3.0** 52.6	**3.0** 52.0	**3.0** 51.8	**3.0** **51.0**
	max	3.0 56.0	3.0 55.0	3.0 58.0	3.0 55.0
SNR=40 dB	min	3.0 50.0	3.0 50.0	3.0 50.0	3.0 50.0
	avg	3.6 60.6	**3.0** 52.4	3.6 87.1	3.3 **51.8**
	max	5.0 77.0	3.0 56.0	4.0 128.0	4.0 55.0
SNR=60 dB	min	3.0 50.0	3.0 50.0	3.0 50.0	3.0 50.0
	avg	3.8 58.2	**3.0** **51.4**	3.8 90.6	3.1 51.9
	max	5.0 86.0	3.0 56.0	5.0 177.0	4.0 55.0

Table 3: All the settings are as in Table 1, except that we report the statistics for both $\widehat{s}$ (top) and $\widehat{n}$ (bottom). We write in bold the average value of $\widehat{s}$ that is closest to $s_{true} = 3$.

Laplacian Noise		ADL/EBIC	ADL_1/EBIC	ADL/BIC	ADL_1/BIC
SNR=20 dB	min	3.0 50.0	3.0 50.0	3.0 50.0	3.0 50.0
	avg	**3.0** **51.3**	**3.0** 51.4	**3.0** 55.3	**3.0** 51.7
	max	3.0 56.0	3.0 55.0	3.0 69.0	3.0 55.0
SNR=40 dB	min	3.0 50.0	3.0 50.0	3.0 59.0	3.0 50.0
	avg	3.7 71.7	**3.0** 52.5	3.8 118.1	3.1 **51.5**
	max	4.0 99.0	3.0 57.0	5.0 193.0	4.0 55.0
SNR=60 dB	min	3.0 50.0	3.0 50.0	3.0 50.0	3.0 50.0
	avg	3.5 76.0	**3.0** 52.6	3.8 110.2	3.1 **51.4**
	max	5.0 123.0	3.0 59.0	5.0 227.0	4.0 57.0

Table 4: All the settings are as in Table 3. Note that the distribution of the additive noise is Laplacian.

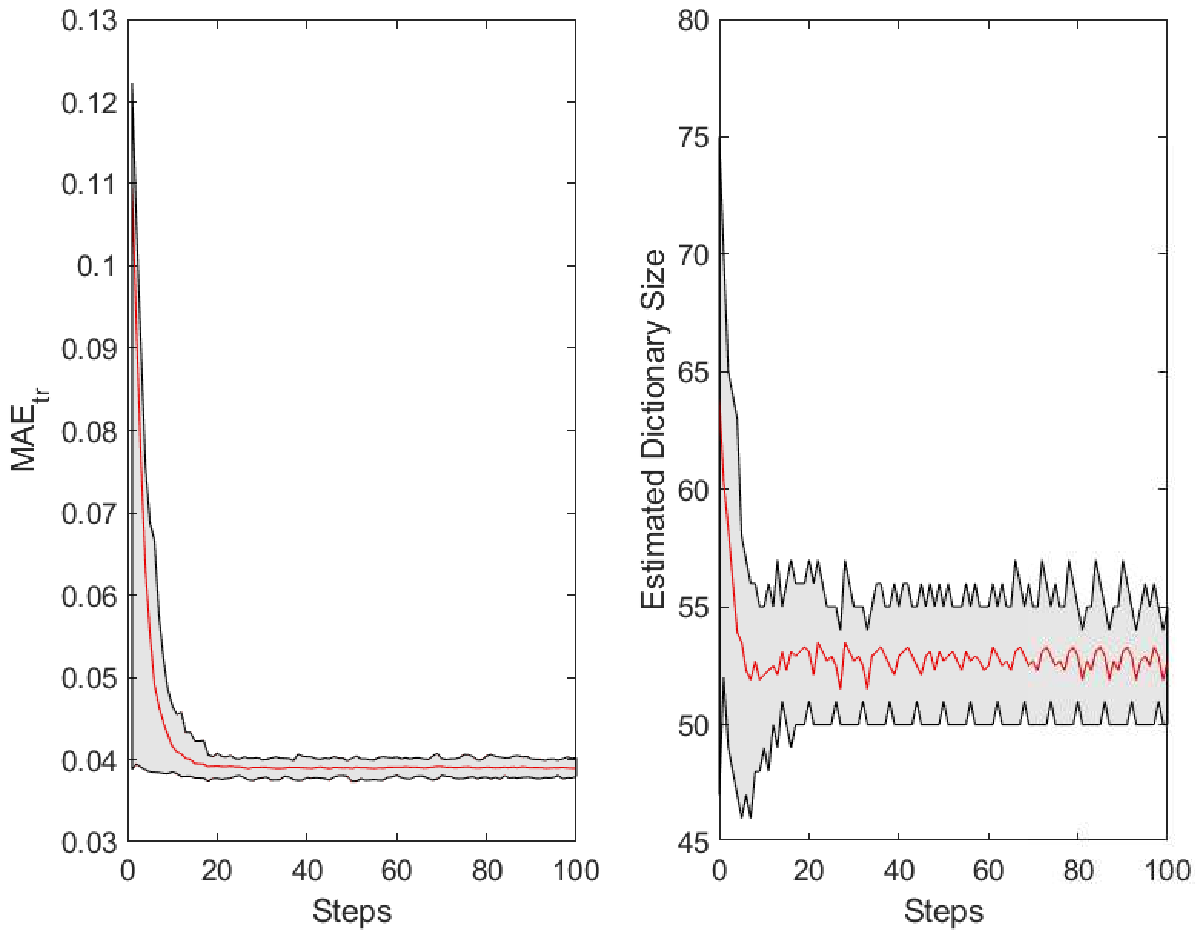

Figure 2: Results obtained by running ADL$_1$/BIC on ten simulated data sets for which the additive noise has Laplacian distribution and SNR=20dB. For each data set, we consider the statistics corresponding to the value $\widehat{s}$ of the sparsity estimated at last step. With the convention that $\widetilde{\boldsymbol{D}}_{step}^{(\widehat{s})}$ and $\widetilde{\boldsymbol{X}}_{step}^{(\widehat{s})}$ are the dictionary and the representation matrix estimated from $\boldsymbol{Y}_{tr}$, the mean absolute error is given by $\mathrm{MAE}_{tr} = \frac{||\boldsymbol{Y}_{tr} - \widetilde{\boldsymbol{D}}_{step}^{(\widehat{s})}\widetilde{\boldsymbol{X}}_{step}^{(\widehat{s})}||_{1,1}}{mN_{tr}}$, where $step \in \{1, \ldots, 100\}$. Left panel: The continuous red line represents the average of MAE_{tr} computed for ten data sets. The upper and the lower bounds of the shaded area correspond to the maximum and to the minimum MAE_{tr} obtained at each step. Right panel: The same conventions as in the left panel, except that the graphical representations are for the dictionary size and not for MAE_{tr}. At $step = 0$, we show the minimum, the average and the maximum of the sizes of dictionaries that are randomly generated for the initialization of the algorithm.

4.3 Comparison with the DL method from [2]

Laplacian Noise		SNR=20dB			
		ADL_1/BIC	MBS-2016/BIC	ADL_1/EBIC	MBS-2016/EBIC
$\widehat{n}$	min	50.0	45.0	50.0	45.0
	avg	51.7	45.0	51.4	47.0
	max	55.0	45.0	55.0	60.0
$\widehat{s}$	min	3.0	1.2	3.0	0.9
	avg	3.0	1.4	3.0	1.3
	max	3.0	1.6	3.0	1.6
ADR(Avg. 10 runs)		1.00	0.51	1.00	0.52
Execution Time (Avg. 10 runs)		102.9secs (s=1) 392.6secs (s=2) 943.5secs (s=3) 1828.0secs (s=4) 3143.6secs (s=5)	1030.5secs (n=45) 1283.7secs (n=50) 1399.0secs (n=55) 1360.8secs (n=60) 1521.6secs (n=65) 1719.4secs (n=70) 1794.9secs (n=75)	100.3secs (s=1) 380.3secs (s=2) 904.4secs (s=3) 1745.2secs (s=4) 2932.1secs (s=5)	1030.5secs (n=45) 1283.7secs (n=50) 1399.0secs (n=55) 1360.8secs (n=60) 1521.6secs (n=65) 1719.4secs (n=70) 1794.9secs (n=75)

Table 5: Comparison with MBS-2016 when SNR=20dB: We report the minimum, the average and the maximum values of $\widehat{n}$ and $\widehat{s}$. Additionally, we present the average ADR computed from ten runs as well as the average execution times.

Laplacian Noise		SNR=40dB			
		ADL_1/BIC	MBS-2016/BIC	ADL_1/EBIC	MBS-2016/EBIC
$\widehat{n}$	min	50.0	45.0	50.0	45.0
	avg	51.5	45.0	52.5	47.5
	max	55.0	45.0	57.0	55.0
$\widehat{s}$	min	3.0	1.4	3.0	1.0
	avg	3.1	1.5	3.0	1.4
	max	4.0	1.6	3.0	1.6
ADR(Avg. 10 runs)		1.00	0.52	1.00	0.56
Execution Time (Avg. 10 runs)		98.0secs (s=1) 365.4secs (s=2) 853.8secs (s=3) 1663.7secs (s=4) 2933.9secs (s=5)	1005.0secs (n=45) 1027.2secs (n=50) 1105.5secs (n=55) 1546.5secs (n=60) 1465.6secs (n=65) 1788.2secs (n=70) 1579.5secs (n=75)	96.1secs (s=1) 360.8secs (s=2) 836.1secs (s=3) 1627.0secs (s=4) 2788.8secs (s=5)	1005.0secs (n=45) 1027.2secs (n=50) 1105.5secs (n=55) 1546.5secs (n=60) 1465.6secs (n=65) 1788.2secs (n=70) 1579.5secs (n=75)

Table 6: The notational conventions are the same as in Table 5, except that SNR=40dB.

Laplacian Noise		SNR=60dB			
		ADL_1/BIC	MBS-2016/BIC	ADL_1/EBIC	MBS-2016/EBIC
$\widehat{n}$	min	50.0	45.0	50.0	45.0
	avg	51.4	45.0	52.6	46.5
	max	57.0	45.0	59.0	55.0
$\widehat{s}$	min	3.0	1.4	3.0	1.1
	avg	3.1	1.5	3.0	1.4
	max	4.0	1.6	3.0	1.5
ADR(Avg. 10 runs)		0.99	0.54	1.00	0.56
Execution Time (Avg. 10 runs)		97.6secs (s=1) 363.6secs (s=2) 844.4secs (s=3) 1666.5secs (s=4) 2968.9secs (s=5)	1102.3secs (n=45) 1274.2secs (n=50) 1262.7secs (n=55) 1391.6secs (n=60) 1541.4secs (n=65) 1531.2secs (n=70) 1803.3secs (n=75)	98.5secs (s=1) 368.0secs (s=2) 845.6secs (s=3) 1644.0secs (s=4) 2860.4secs (s=5)	1102.3secs (n=45) 1274.2secs (n=50) 1262.7secs (n=55) 1391.6secs (n=60) 1541.4secs (n=65) 1531.2secs (n=70) 1803.3secs (n=75)

Table 7: The notational conventions are the same as in Table 5, except that SNR=60dB.

4.4 Comparison with the DL method based on [3–5]

Laplacian Noise		SNR=20dB			
		ADL_1/BIC	LW-2019/BIC	ADL_1/EBIC	LW-2019/EBIC
$\widehat{n}$	min	50.0	50.0	50.0	45.0
	avg	51.7	56.0	51.4	47.0
	max	55.0	65.0	55.0	50.0
$\widehat{s}$	min	3.0	3.0	3.0	1.0
	avg	3.0	3.0	3.0	1.0
	max	3.0	3.0	3.0	1.0
ADR(Avg. 10 runs)		1.00	0.97	1.00	0.20
Execution Time (Avg. 10 runs)		102.9secs (s=1)	535.3secs (s=1)	100.3secs (s=1)	535.3secs (s=1)
		392.6secs (s=2)	1063.0secs (s=2)	380.3secs (s=2)	1063.0secs (s=2)
		943.5secs (s=3)	1689.2secs (s=3)	904.4secs (s=3)	1689.2secs (s=3)
		1828.0secs (s=4)	2071.0secs (s=4)	1745.2secs (s=4)	2071.0secs (s=4)
		3143.6secs (s=5)	2983.4secs (s=5)	2932.1secs (s=5)	2983.4secs (s=5)

Table 8: Comparison with LW-2019 when SNR=20dB: As in Table 5, we report the minimum, the average and the maximum values of $\widehat{n}$ and $\widehat{s}$. Additionally, we present the average ADR computed from ten runs as well as the average execution times.

Laplacian Noise		SNR=40dB			
		ADL_1/BIC	LW-2019/BIC	ADL_1/EBIC	LW-2019/EBIC
$\widehat{n}$	min	50.0	50.0	50.0	50.0
	avg	51.5	59.0	52.5	59.0
	max	55.0	75.0	57.0	75.0
$\widehat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.1	3.0	3.0	3.0
	max	4.0	3.0	3.0	3.0
ADR(Avg. 10 runs)		1.00	1.00	1.00	1.00
Execution Time (Avg. 10 runs)		98.0secs (s=1)	548.7secs (s=1)	96.1secs (s=1)	548.7secs (s=1)
		365.4secs (s=2)	1099.0secs (s=2)	360.8secs (s=2)	1099.0secs (s=2)
		853.8secs (s=3)	1640.1secs (s=3)	836.1secs (s=3)	1640.1secs (s=3)
		1663.7secs (s=4)	2192.9secs (s=4)	1627.0secs (s=4)	2192.9secs (s=4)
		2933.9secs (s=5)	3007.5secs (s=5)	2788.8secs (s=5)	3007.5secs (s=5)

Table 9: The notational conventions are the same as in Table 8, except that SNR=40dB.

Laplacian Noise		SNR=60dB			
		ADL_1/BIC	LW-2019/BIC	ADL_1/EBIC	LW-2019/EBIC
$\widehat{n}$	min	50.0	50.0	50.0	50.0
	avg	51.4	58.5	52.6	58.5
	max	57.0	70.0	59.0	70.0
$\widehat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.1	3.0	3.0	3.0
	max	4.0	3.0	3.0	3.0
ADR(Avg. 10 runs)		0.99	1.00	1.00	1.00
Execution Time (Avg. 10 runs)		97.6secs (s=1) 363.6secs (s=2) 844.4secs (s=3) 1666.5secs (s=4) 2968.9secs (s=5)	534.1secs (s=1) 1058.0secs (s=2) 1684.3secs (s=3) 2276.7secs (s=4) 2949.7secs (s=5)	98.5secs (s=1) 368.0secs (s=2) 845.6secs (s=3) 1644.0secs (s=4) 2860.4secs (s=5)	534.1secs (s=1) 1058.0secs (s=2) 1684.3secs (s=3) 2276.7secs (s=4) 2949.7secs (s=5)

Table 10: The notational conventions are the same as in Table 8, except that SNR=60dB.

4.5 Comparison with the DL method from [6]

4.5.1 Data simulated with Laplacian additive noise

Laplacian Noise	SNR = 20 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.81	0.1	0.2	0.3	0.00	0.06	0.15
n =50	0.94	0.1	0.2	0.3	0.00	0.05	0.15
n =55	0.97	0.1	0.2	0.4	0.00	0.04	0.15
n =60	0.97	0.1	0.2	0.3	0.00	0.03	0.10
n =65	0.96	0.1	0.1	0.2	0.00	0.03	0.10
n =70	0.94	0.1	0.1	0.2	0.00	0.05	0.15
n =75	0.94	0.1	0.1	0.1	0.00	0.04	0.10

Table 11: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 4 (for SNR=20dB).

Laplacian Noise	SNR = 40 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.82	0.1	0.2	0.4	0.00	0.07	0.15
n =50	0.96	0.1	0.2	0.3	0.00	0.04	0.10
n =55	0.99	0.1	0.2	0.2	0.00	0.05	0.15
n =60	0.98	0.1	0.1	0.3	0.00	0.04	0.15
n =65	0.98	0.1	0.1	0.2	0.00	0.04	0.10
n =70	0.98	0.1	0.1	0.2	0.00	0.06	0.15
n =75	0.96	0.1	0.1	0.1	0.00	0.03	0.10

Table 12: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 4 (for SNR=40dB).

Laplacian Noise	SNR = 60 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.81	0.1	0.2	0.3	0.00	0.05	0.15
n =50	0.95	0.1	0.2	0.3	0.00	0.05	0.15
n =55	0.98	0.1	0.2	0.4	0.00	0.04	0.10
n =60	0.98	0.1	0.1	0.3	0.00	0.04	0.10
n =65	0.97	0.1	0.1	0.2	0.00	0.04	0.15
n =70	0.98	0.1	0.1	0.1	0.00	0.03	0.10
n =75	0.97	0.1	0.1	0.2	0.00	0.04	0.10

Table 13: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 4 (for SNR=60dB).

4.5.2 Data simulated with GMM additive noise

GMM Noise		SNR=20dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	50.0	50.0	50.0	50.0
	avg	51.9	72.4	50.8	51.5
	max	56.0	178.0	56.0	55.0
$\hat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.0	3.3	3.0	3.0
	max	3.0	5.0	3.0	3.0
ADR(Avg. 10 runs)		1.00	1.00	1.00	1.00

Table 14: GMM additive noise: Statistics collected for ADL and ADL_1 from 10 trials. Note that $n_{true} = 50$ and $s_{true} = 3$.

GMM Noise		SNR=40dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	50.0	51.0	50.0	50.0
	avg	51.7	93.0	50.6	61.9
	max	55.0	156.0	55.0	101.0
$\hat{s}$	min	3.0	4.0	3.0	3.0
	avg	3.0	4.3	3.0	3.8
	max	3.0	5.0	3.0	5.0
ADR(Avg. 10 runs)		1.00	0.98	1.00	0.97

Table 15: The notational conventions are the same as in Table 14, except that SNR=40dB.

GMM Noise		SNR=60dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	50.0	60.0	50.0	50.0
	avg	51.1	99.8	50.9	65.7
	max	55.0	146.0	58.0	110.0
$\hat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.0	3.8	3.0	3.6
	max	3.0	5.0	3.0	5.0
ADR(Avg. 10 runs)		1.00	0.98	1.00	0.98

Table 16: The notational conventions are the same as in Table 14, except that SNR=60dB.

GMM Noise	SNR = 20 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.81	0.1	0.2	0.3	0.00	0.04	0.10
n =50	0.94	0.1	0.2	0.3	0.00	0.04	0.20
n =55	0.99	0.1	0.2	0.3	0.00	0.04	0.10
n =60	0.97	0.1	0.1	0.3	0.00	0.05	0.15
n =65	0.97	0.1	0.1	0.2	0.00	0.04	0.10
n =70	0.95	0.1	0.1	0.4	0.00	0.03	0.10
n =75	0.94	0.1	0.1	0.3	0.00	0.04	0.10

Table 17: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 14.

GMM Noise	SNR = 40 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.82	0.1	0.2	0.3	0.00	0.04	0.10
n =50	0.96	0.1	0.2	0.4	0.00	0.05	0.15
n =55	0.98	0.1	0.2	0.3	0.00	0.04	0.10
n =60	0.98	0.1	0.1	0.2	0.00	0.05	0.15
n =65	0.98	0.1	0.1	0.2	0.00	0.05	0.15
n =70	0.99	0.1	0.1	0.2	0.00	0.04	0.10
n =75	0.96	0.1	0.1	0.2	0.00	0.05	0.15

Table 18: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 15.

GMM Noise	SNR = 60 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.82	0.1	0.2	0.3	0.00	0.05	0.15
n =50	0.97	0.1	0.2	0.2	0.00	0.06	0.10
n =55	0.98	0.1	0.1	0.3	0.00	0.05	0.15
n =60	0.99	0.1	0.1	0.2	0.00	0.05	0.15
n =65	0.99	0.1	0.1	0.2	0.00	0.04	0.10
n =70	0.96	0.1	0.1	0.2	0.00	0.05	0.15
n =75	0.96	0.1	0.1	0.2	0.00	0.04	0.10

Table 19: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 16.

GMM Noise	SNR = 20 dB			
	λ_{min}		λ_{max}	
	δ_{min}	δ_{max}	δ_{min}	δ_{max}
$n = 45$	5.94 secs	5.93 secs	2.88 secs	3.41 secs
$n = 75$	7.85 secs	9.35 secs	4.48 secs	5.70 secs

Table 20: Execution time for IS-2019: All values are calculated as an average over ten runs. We report results only for the smallest dictionary size ($n = 45$) and for the largest dictionary size ($n = 75$). For both dictionary sizes, we have $\lambda_{min} = 0.1$, $\lambda_{max} = 0.3$, $\delta_{min} = 0.00$ and $\delta_{max} = 0.10$ (see Table 17). Note that the execution times in the table are computed for one pair of parameters (λ, δ). For obtaining the ADR values shown in Table 17, we have run IS-2019 for 561 $(11 \cdot 51)$ different pairs (λ, δ).

4.5.3 Data simulated with Cauchy additive noise

The samples of the additive noise are drawn from a Cauchy distribution for which the location parameter is zero and the scale parameter is $\sigma > 0$. Due to the particularities of the Cauchy distribution, SNR is computed with formula from [10]: $\text{SNR} = 10\log_{10}(1/\sigma)$. The results shown in Tables 21-23 demonstrate again that ADL_1 is superior to ADL. When SNR is low, ADL_1 has difficulties in one single trial, where the size of the dictionary is severely underestimated by both BIC and EBIC. However, the ADR computed for other trials is much better. In the case of ADL_1/BIC, ADR = 0.92 (for one trial), ADR = 0.98 (for two trials) and ADR = 1.00 (for six trials). If SNR is greater or equal to 40dB, then ADR equals 1.00 in all trials. For comparison, we display in Tables 24-26 the results produced by IS-2019 on the same data sets. They are similar to those obtained for Laplacian and GMM noise, in the sense that almost all the atoms of the true dictionary are recovered when the dictionary size is large enough, but IS-2019 does not yield an average ADR of 1.00. It is interesting to note that the values of δ in Table 24 are greater than in Table 17. Hence, IS-2019 is able to detect that in the second scenario there are more outliers than in the first one. This is achieved by using the information about the ground truth when searching in the parameter space.

Cauchy Noise		SNR=20dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	20.0	81.0	20.0	74.0
	avg	64.7	94.1	64.3	79.3
	max	72.0	124.0	73.0	87.0
$\hat{s}$	min	1.0	3.0	1.0	3.0
	avg	3.1	3.9	2.9	3.3
	max	4.0	4.0	4.0	4.0
ADR(Avg. 10 runs)		0.90	1.00	0.90	1.00

Table 21: Cauchy additive noise: Statistics collected for ADL and ADL_1 from 10 trials. Note that $n_{true} = 50$ and $s_{true} = 3$.

Cauchy Noise		SNR=40dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	50.0	59.0	50.0	50.0
	avg	52.0	132.9	51.3	74.4
	max	57.0	229.0	57.0	133.0
$\hat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.0	4.0	3.0	3.2
	max	3.0	5.0	3.0	4.0
ADR(Avg. 10 runs)		1.00	0.99	1.00	1.00

Table 22: The notational conventions are the same as in Table 21, except that SNR=40dB.

Cauchy Noise		SNR=60dB			
		ADL_1/BIC	ADL/BIC	ADL_1/EBIC	ADL/EBIC
$\hat{n}$	min	50.0	50.0	50.0	50.0
	avg	51.6	81.1	51.2	59.6
	max	56.0	174.0	56.0	101.0
$\hat{s}$	min	3.0	3.0	3.0	3.0
	avg	3.0	3.7	3.0	3.8
	max	3.0	5.0	3.0	5.0
ADR(Avg. 10 runs)		1.00	0.98	1.00	0.96

Table 23: The notational conventions are the same as in Table 21, except that SNR=60dB.

Cauchy Noise	SNR = 20 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.75	0.2	0.2	0.3	0.06	0.11	0.15
n =50	0.88	0.2	0.2	0.3	0.06	0.11	0.20
n =55	0.95	0.2	0.2	0.3	0.06	0.10	0.15
n =60	0.98	0.2	0.2	0.3	0.06	0.09	0.15
n =65	0.96	0.1	0.2	0.3	0.06	0.08	0.15
n =70	0.96	0.1	0.2	0.3	0.00	0.09	0.15
n =75	0.95	0.1	0.2	0.4	0.00	0.06	0.10

Table 24: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 21.

Cauchy Noise	SNR = 40 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.80	0.1	0.2	0.4	0.00	0.06	0.15
n =50	0.94	0.1	0.2	0.4	0.00	0.06	0.15
n =55	0.98	0.1	0.1	0.3	0.00	0.05	0.15
n =60	0.98	0.1	0.1	0.3	0.00	0.05	0.15
n =65	0.98	0.1	0.1	0.3	0.00	0.03	0.10
n =70	0.99	0.1	0.1	0.1	0.00	0.04	0.10
n =75	0.96	0.1	0.1	0.2	0.00	0.04	0.10

Table 25: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 22.

Cauchy Noise	SNR = 60 dB						
	ADR	λ			δ		
		min	avg	max	min	avg	max
n =45	0.82	0.1	0.2	0.3	0.00	0.04	0.15
n =50	0.95	0.1	0.2	0.4	0.00	0.04	0.15
n =55	0.98	0.1	0.1	0.2	0.00	0.05	0.15
n =60	0.98	0.1	0.1	0.2	0.00	0.04	0.10
n =65	0.98	0.1	0.1	0.2	0.00	0.04	0.10
n =70	0.97	0.1	0.1	0.1	0.00	0.05	0.10
n =75	0.97	0.1	0.1	0.1	0.00	0.04	0.10

Table 26: Statistics collected for IS-2019 when the simulated data sets are the same as those used to produce the results reported in Table 23.

4.6 Shape parameter estimation

Data set	ExpShape1/MAE	ExpShape2/MAE	ExpShape1/EBIC	ExpShape2/EBIC
#1	1.3119	1.4305	1.3119	1.5492
#2	2.2373	1.5017	1.2644	1.3831
#3	1.1932	1.3593	1.0983	2.0949
#4	1.3593	1.4305	1.3119	1.5492
#5	1.2644	1.4542	1.3831	1.3593
#6	1.2881	1.3831	1.4305	1.5492
#7	1.3831	1.6678	1.5966	1.6678
#8	2.1424	1.3593	2.1186	1.5729
#9	1.5254	1.4305	1.5254	1.6441
#10	1.2881	1.4068	1.2881	1.4542
Mean	1.4993	1.4424	1.4329	1.5824
Variance	0.1406	0.0082	0.0776	0.0426

Table 27: Experiments ExpShape1 and ExpShape2: For each simulated data set, we show the values estimated for the shape parameter when MAE (computed on the validation set) and EBIC (computed on the training set) are employed. In each case, the mean and the variance of the estimated values are reported. Note that $p_{true} = 1.5$.

5 Experiments with signals generated from images

5.1 Results obtained with JPEG data

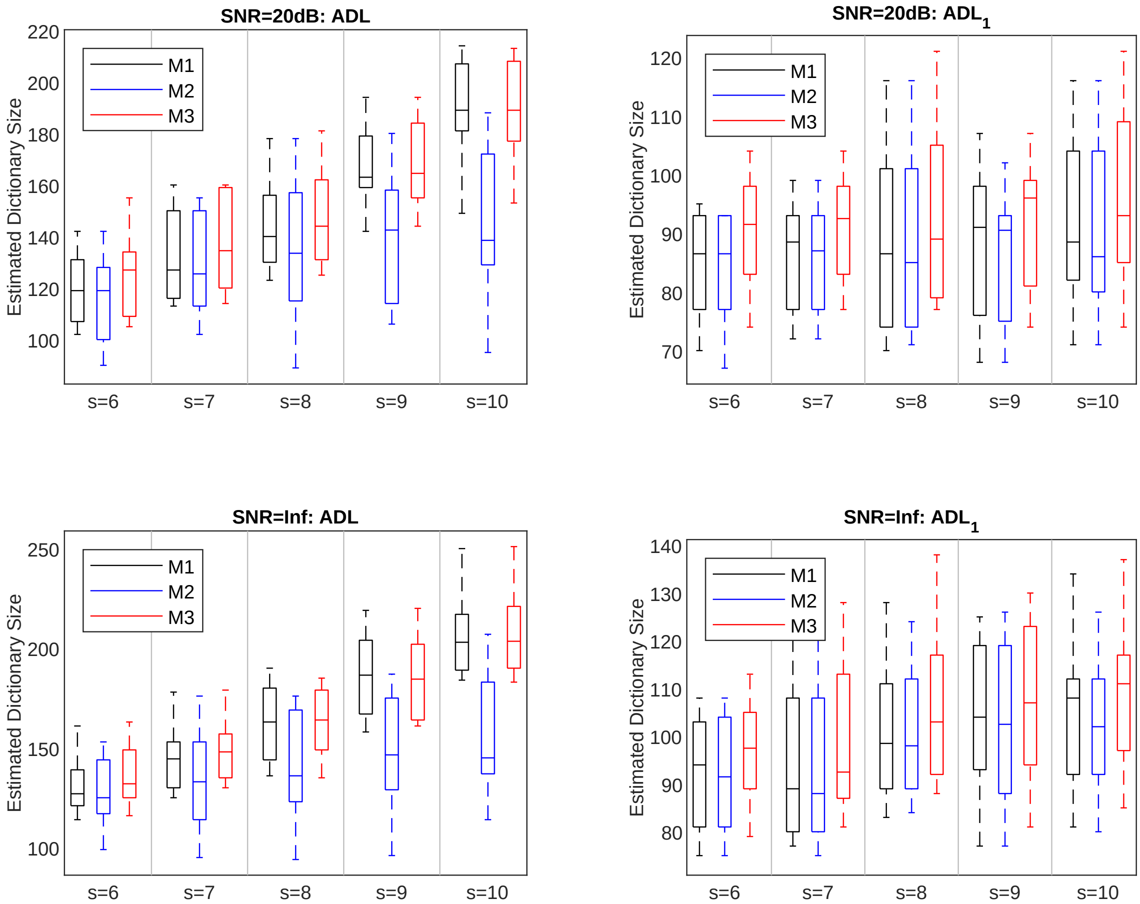

Figure 3: JPEG data: Sizes of the dictionaries estimated by the methods M1, M2 and M3 when the values of s are those written on the horizontal axes. The value of the SNR and the name of the algorithm are mentioned in the title of each panel. The results are collected from 10 data sets.

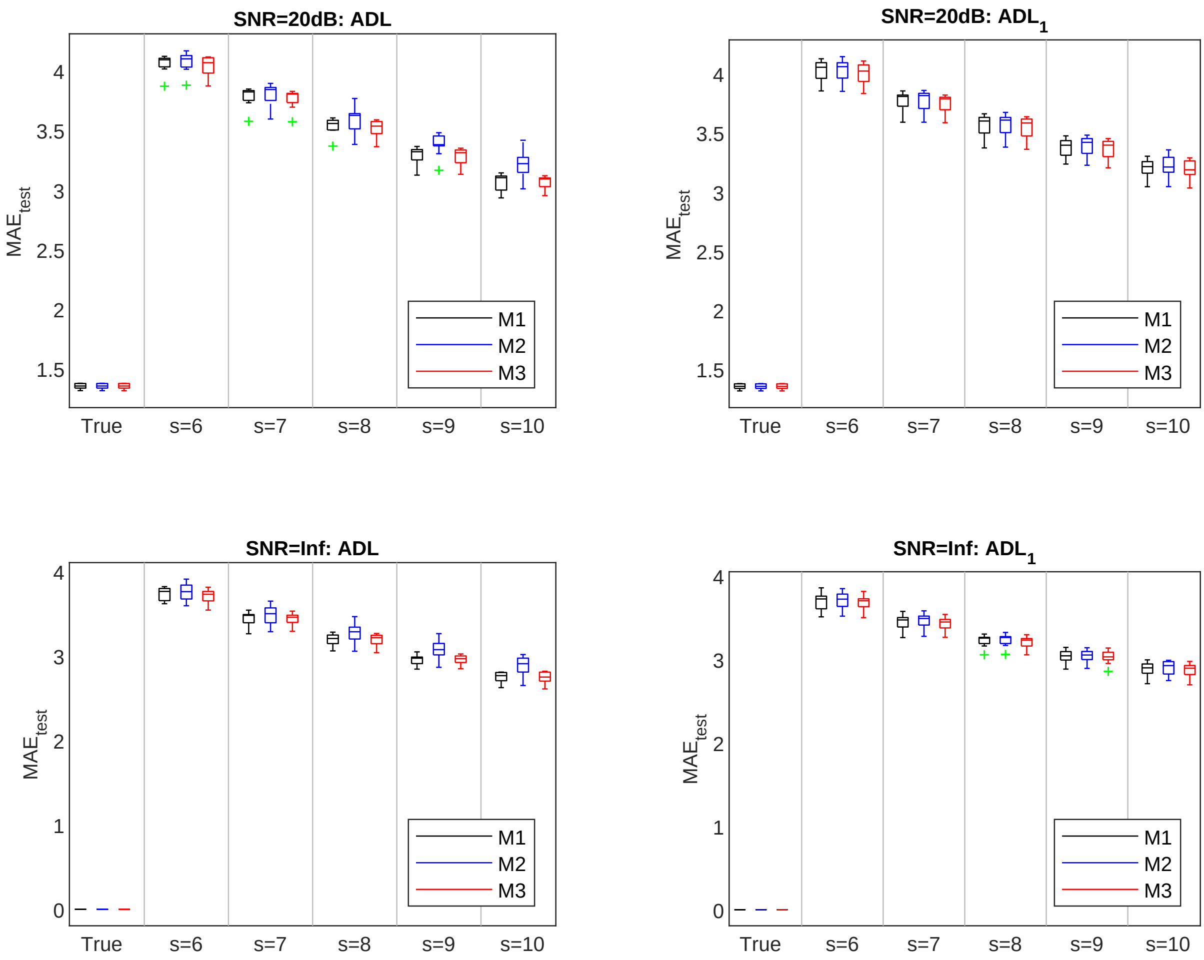

Figure 4: JPEG data: Values of MAE_{test} obtained by three methods. For comparison, we show in the column labeled "True" the level of the noise for the test set. This is calculated by replacing in [1, Eq. (32)] the product $\widehat{\boldsymbol{D}}\widehat{\boldsymbol{X}}_{test}$ with the noiseless data. All graphical conventions are the same as in Figure 3.

5.2 Results obtained with MRI data

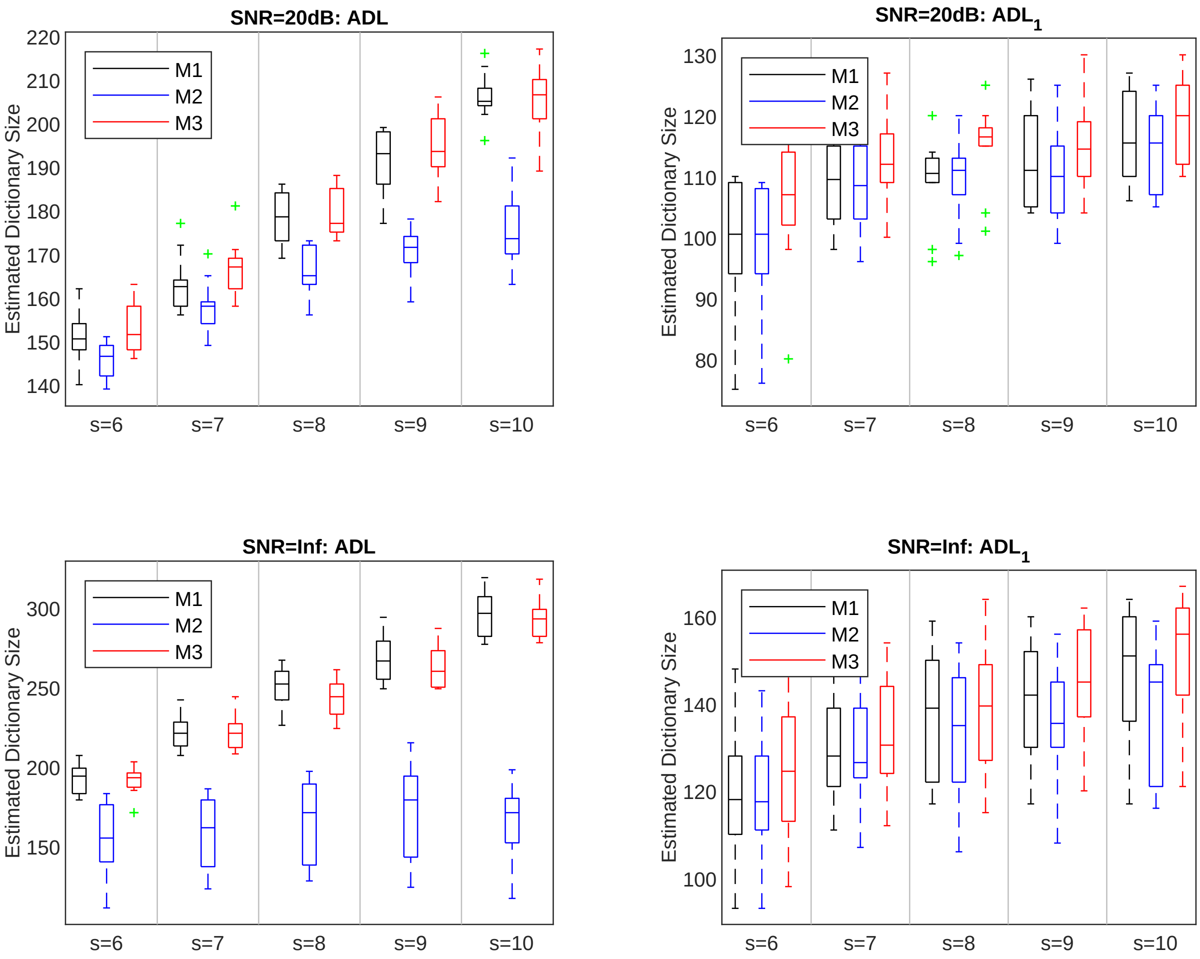

Figure 5: MRI data: Same settings as in Figure 3.

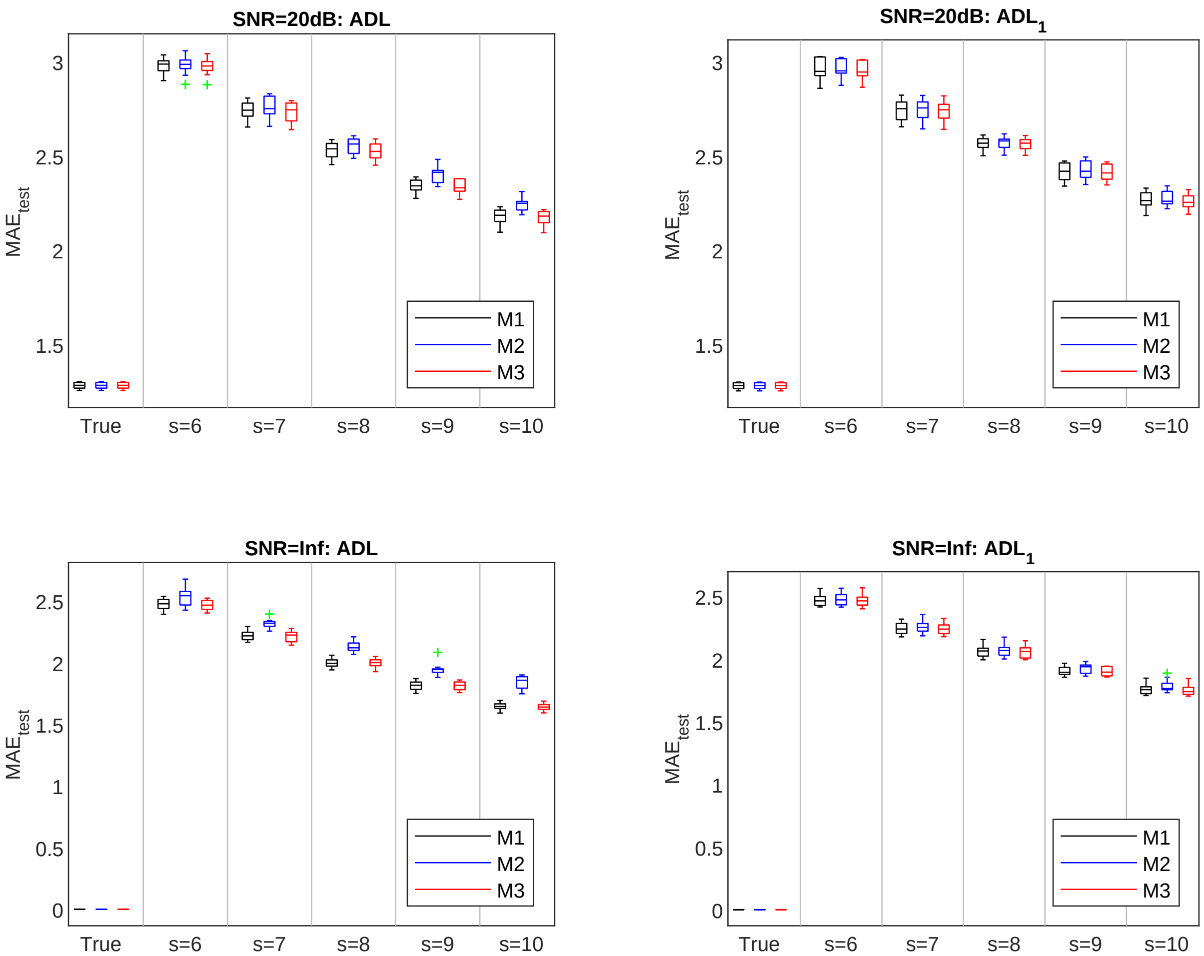

Figure 6: MRI data: Same settings as in Figure 4.

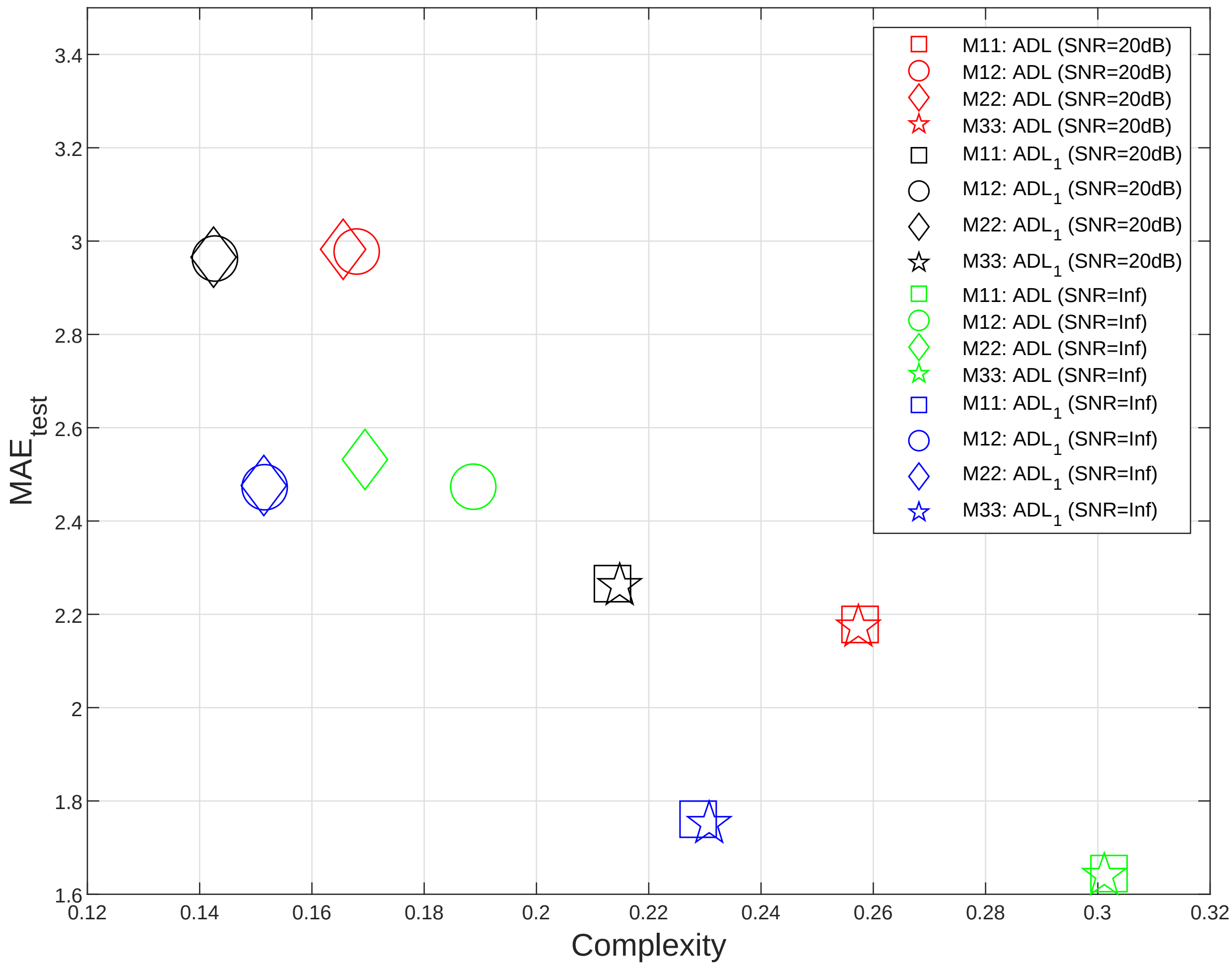

Figure 7: MRI data: Test errors versus the complexity of the models selected during training. In each case, MAE_{test} and Complexity are calculated as averages for ten data sets. See also [1, Fig. 3].

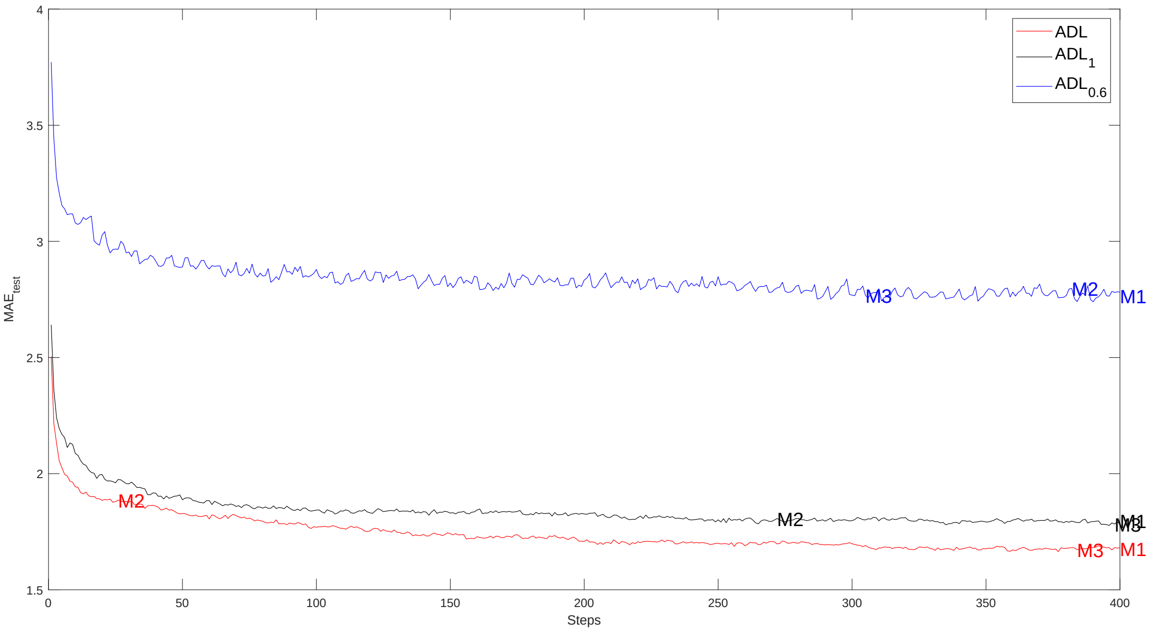

Figure 8: The values of MAE_{test} at each step, when ADL, ADL_1 and $ADL_{0.6}$ are applied (with $s = 10$). See also [1, Fig. 4 (left panel)].

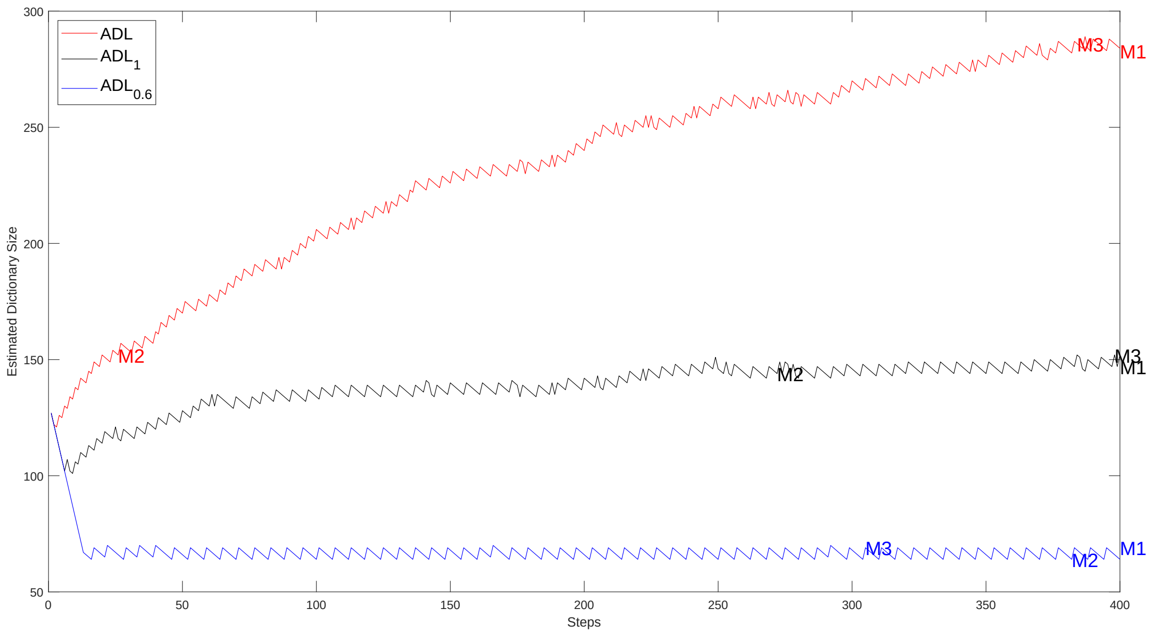

Figure 9: The sizes of the dictionaries at each step, when ADL, ADL_1 and $ADL_{0.6}$ are applied (with $s = 10$). See also [1, Fig. 4 (right panel)].

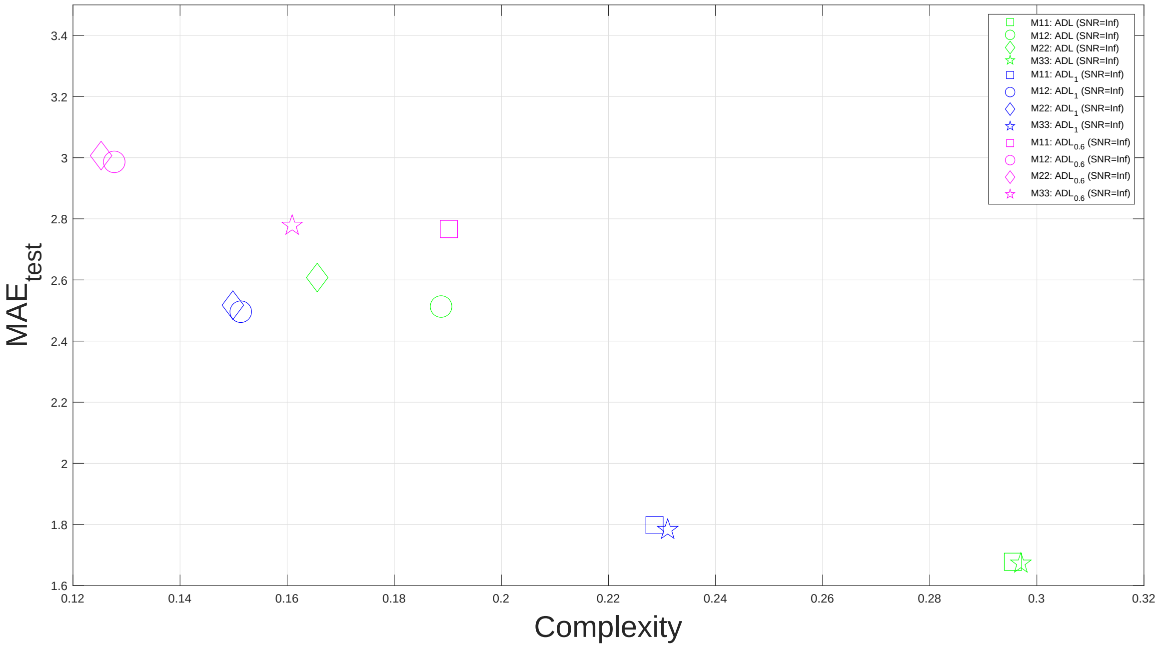

Figure 10: Same conventions as in Fig. 7, except that the results for each method are generated from the first noiseless MRI data set.

References

[1] X. Zheng, B. Dumitrescu, J. Liu, C. Giurcăneanu, Dictionary learning for signals in additive noise with generalized Gaussian distribution, 2022.

[2] S. Mukherjee, R. Basu, C. S. Seelamantula, L1-K-SVD: A robust dictionary learning algorithm with simultaneous update, Signal Processing 123 (2016) 42–52.

[3] C. A. Loza, Robust variants of dictionary learning exploiting M-estimators, in: 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), 2019, pp. 1–6. doi:10.1109/CHILECON47746.2019.8988048.

[4] C. A. Loza, RobOMP: Robust variants of orthogonal matching pursuit for sparse representations, PeerJ Computer Science (2019) 5:e192.

[5] N. Kwak, Principal component analysis based on l1-norm maximization, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2008) 1672–1680.

[6] A. Iqbal, A.-K. Seghouane, An α-divergence-based approach for robust dictionary learning, IEEE Transactions on Image Processing 28 (2019) 5729–5739.

[7] P. Stoica, P. Babu, On the proper forms of BIC for model order selection, IEEE Transactions on Signal Processing 60 (2012) 4956–4961.

[8] S. Kay, Fundamentals of statistical signal processing: estimation theory, Prentice Hall, 1993.

[9] M. DeGroot, M. Schervish, Probability and statistics, 4th ed., Pearson Education, Inc., 2012.

[10] W.-J. Zeng, H. So, X. Jiang, Outlier-robust greedy pursuit algorithms in ℓ_p -space for sparse approximation, IEEE Transactions on Signal Processing 64 (2016) 60–75.