SCIENCE CHINA
Information Sciences

• LETTER •

February 2023, Vol. 66 129202:1–129202:2
https://doi.org/10.1007/s11432-022-3546-5

A recursive least squares algorithm with ℓ_1 regularization for sparse representation

Di LIU[1,2], Simone BALDI[1,3*], Quan LIU[4] & Wenwu YU[1,3*]

[1]*School of Cyber Science and Engineering, Southeast University, Nanjing* 210096*, China;*
[2]*School of Computation, Information and Technology, Technical University of Munich, Garching* 85748*, Germany;*
[3]*School of Mathematics, Center for Mobile Communication and Security, Southeast University, Nanjing* 210096*, China;*
[4]*School of Artificial Intelligence, Southeast University, Suzhou* 215100*, China*

Received 9 March 2022/Revised 8 May 2022/Accepted 17 June 2022/Published online 5 January 2023

Citation Liu D, Baldi S, Liu Q, et al. A recursive least squares algorithm with ℓ_1 regularization for sparse representation. Sci China Inf Sci, 2023, 66(2): 129202, https://doi.org/10.1007/s11432-022-3546-5

Sparse representation aims to identify a few basic elements in a signal, so as to use a combination of such elements to reconstruct the original signal. The ℓ_1-norm has been widely applied in sparse representation, either when processing batches of data offline, or online as in adaptive filtering [1–3]. Representative algorithms for online sparse representation include zero-attracting LMS (ZA-LMS) [4], ℓ_1-RLS [5], ℓ_1-RRLS [6], zero-attracting RLS (ZA-RLS) [7], among others [8, 9]. A new recursive least squares (RLS) method for sparse representation, named ℓ_1^2-RLS, is proposed in this study. The method, relying on minimizing an appropriately designed cost with ℓ_1 regularization, is compared with state-of-the-art algorithms adopting ℓ_2 and ℓ_1 regularization.

Proposed ℓ_1^2-RLS method. Consider an input-output parameter estimation setting given by the standard relation:

$$y(k) = \sum_{i=0}^{N-1} h_i x_i(k) + n(k) = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}(k) + n(k), \qquad (1)$$

where $\boldsymbol{h} = [h_0\, h_1 \cdots h_{N-1}]^{\mathrm{T}}$ denotes the unknown weight vector to be estimated, $\boldsymbol{x}(k) = [x_0(k)\, x_1(k) \cdots x_{N-1}(k)]^{\mathrm{T}}$ is the input signal, $y(k)$ is the output signal, $n(k)$ is the measurement noise, and k is the time index. The system is sparse when only a few elements of $\boldsymbol{h}$ are non-zero. The estimation goal is to provide an estimate of $\boldsymbol{h}$ at time k, called $\widehat{\boldsymbol{h}}(k)$, by using input and output signals collected up to time k.

To formulate the estimation problem, let us define the following least squares cost function with ℓ_1-regularization:

$$\begin{aligned} J_N(k) \;=\; & \frac{1}{2}\sum_{i=1}^{k}\left(y(i) - \widehat{\boldsymbol{h}}^{\mathrm{T}}(k)\boldsymbol{x}(i)\right)^2 + \frac{\rho}{2}||\widehat{\boldsymbol{h}}(k)||_1^2 \qquad (2)\\ & + \frac{1}{2}(\widehat{\boldsymbol{h}}(k) - \widehat{\boldsymbol{h}}^{\mathrm{T}}(0))P^{-1}(0)(\widehat{\boldsymbol{h}}(k) - \widehat{\boldsymbol{h}}(0)), \end{aligned}$$

where $\widehat{\boldsymbol{h}}(0)$ is an initial estimate of $\boldsymbol{h}$, $P(0)$ is an initial covariance matrix that weights the ℓ_2-norm, and $\rho > 0$ is the regularizing parameter that weights the ℓ_1-norm. Note the square of the ℓ_1-norm in (2), consistent with the square of the ℓ_2-norm in Tikhonov regularization. The resulting ℓ_1^2-RLS method is in Algorithm 1. The details of the derivation of the algorithm are in Appendix A.

Algorithm 1 Proposed ℓ_1^2 regularized recursive least squares (ℓ_1^2-RLS)

Data and parameters: $\boldsymbol{x}(n), y(n)$, ρ, $\widehat{\boldsymbol{h}}(0)$, $P(0)$.
1: **for** time step $k = 1, 2, \ldots, n$, **do**
2: Let $e(k) = y(k) - \widehat{\boldsymbol{h}}^{\mathrm{T}}(k-1)\boldsymbol{x}(k)$ and $\overline{\mathbf{sgn}}(k-1) = [\mathrm{sgn}(\widehat{h}_0(k-1))\ \mathrm{sgn}(\widehat{h}_1(k-1)) \cdots \mathrm{sgn}(\widehat{h}_{N-1}(k-1))]^{\mathrm{T}}$;
3: Let $Q(k) = [\boldsymbol{x}(k)\ \ \sqrt{\rho}\overline{\mathbf{sgn}}(k-1)\ \ -\sqrt{\rho}\overline{\mathbf{sgn}}(k-2)]$;
4: Let $S(k) = [\boldsymbol{x}(k)\ \ \sqrt{\rho}\,\overline{\mathbf{sgn}}(k-1)\ \ \sqrt{\rho}\,\overline{\mathbf{sgn}}(k-2)]^{\mathrm{T}}$;
5: Update $P(k) = P(k-1) - P(k-1)Q(k) \times (I + S(k)P(k-1)Q(k))^{-1}S(k)P(k-1)$;
6: Update $\widehat{\boldsymbol{h}}(k) = \widehat{\boldsymbol{h}}(k-1) + P(k)\boldsymbol{x}(k)e(k)\ -\rho P(k)[\overline{\mathbf{sgn}}(k-1)\ -\overline{\mathbf{sgn}}(k-2)]\begin{bmatrix}\overline{\mathbf{sgn}}^{\mathrm{T}}(k-1)\\ \overline{\mathbf{sgn}}^{\mathrm{T}}(k-2)\end{bmatrix}\widehat{\boldsymbol{h}}(k-1)$;
7: **end for**

Remark 1. By looking at steps 5 and 6 of Algorithm 1, one can notice that the proposed ℓ_1^2-RLS algorithm adds a few extra terms to the standard RLS algorithm, which are

$$\begin{aligned} P(k) \;&=\; P(k-1) - \frac{P(k-1)\boldsymbol{x}(k)\boldsymbol{x}^{\mathrm{T}}(k)P(k-1)}{1 + \boldsymbol{x}^{\mathrm{T}}(k)P(k-1)\boldsymbol{x}(k)},\\ \widehat{\boldsymbol{h}}(k) \;&=\; \widehat{\boldsymbol{h}}(k-1) + P(k)\boldsymbol{x}(k)e(k). \end{aligned}$$

In step 5 of Algorithm 1, new terms appear resulting from the matrix inversion lemma: the inverse in step 5 involves a 3×3 matrix $I + S(k)P(k-1)Q(k)$ instead of a scalar inverse as in the standard RLS. In step 6, notice that the term

$$-\rho P(k)[\overline{\mathbf{sgn}}(k-1)\ \ -\overline{\mathbf{sgn}}(k-2)]\begin{bmatrix}\overline{\mathbf{sgn}}^{\mathrm{T}}(k-1)\\ \overline{\mathbf{sgn}}^{\mathrm{T}}(k-2)\end{bmatrix}\widehat{\boldsymbol{h}}(k-1)$$

plays the role of attracting the estimate $\widehat{\boldsymbol{h}}(k)$ towards zero. In this sense, the philosophy to induce sparsity is analogous

* Corresponding author (email: s.baldi@tudelft.nl, wwyu@seu.edu.cn)

Table 1 Summary of comparisons (the best performance is highlighted with bold notation)

	Effect of regularization, $K=3$						Effect of sparsity				
	$\rho=0.1$	$\rho=0.5$	$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=5$	K=1	K=3	K=5	K=7	K=9
RLS	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400
ℓ_1-RRLS	0.0401	0.0400	0.0399	0.0398	0.0397	0.0391	0.0397	0.0399	0.0400	0.0401	0.0401
ZA-RLS	0.0399	0.0397	0.0393	0.0390	0.0387	0.0370	**0.0392**	0.0393	0.0394	0.0397	0.0399
ℓ_1^2-RLS (proposed)	**0.0398**	**0.0396**	**0.0392**	**0.0389**	**0.0386**	**0.0368**	0.0396	**0.0392**	**0.0391**	**0.0393**	**0.0397**

	Effect of regularization, $K=5$						Effect of signal-to-noise ratio				
	$\rho=0.1$	$\rho=0.5$	$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=5$	SNR=1	SNR=3	SNR=5	SNR=7	SNR=10
RLS	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400	0.0503	0.0400	0.0317	0.0252	0.0179
ℓ_1-RRLS	0.0401	0.0401	0.0400	0.0399	0.0399	0.0395	0.0503	0.0399	0.0316	0.0252	0.0177
ZA-RLS	0.0399	0.0397	0.0394	0.0392	0.0390	0.0378	0.0497	0.0393	0.0311	0.0249	**0.0172**
ℓ_1^2-RLS (proposed)	**0.0398**	**0.0395**	**0.0391**	**0.0388**	**0.0384**	**0.0368**	**0.0496**	**0.0392**	**0.0310**	**0.0248**	**0.0172**

	Effect of regularization, $K=7$					
	$\rho=0.1$	$\rho=0.5$	$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=5$
RLS	0.0400	0.0400	0.0400	0.0400	0.0400	0.0400
ℓ_1-RRLS	0.0401	0.0401	0.0401	0.0400	0.0400	0.0398
ZA-RLS	0.0399	0.0398	0.0397	0.0395	0.0394	0.0388
ℓ_1^2-RLS (proposed)	**0.0398**	**0.0396**	**0.0393**	**0.0391**	**0.0389**	**0.0385**

to the zero-attracting methods [4, 7]. When $\rho = 0$, steps 5 and 6 in Algorithm 1 degenerate to the standard RLS.

Comparative experiments. The performance of the proposed ℓ_1^2-RLS is compared with the standard RLS [1], ℓ_1-RRLS [6], and ZA-RLS [7]. The input $\boldsymbol{x}(k)$ is white and $n(k)$ is additive white Gaussian noise with a certain signal-to-noise ratio (SNR). The system is as in (1) with a total of 10 coefficients, where only K of them are nonzero. For every experiment, $\boldsymbol{h}$ is normalized in so that $||\boldsymbol{h}||_1 = ||\boldsymbol{h}||_1^2 = 1$, which guarantees fair numerical comparisons as the penalty with ρ has similar effect for all algorithms. The performance of all methods is tested in three aspects.

(1) The effect of regularization parameter ρ on the performance. This is done because increasing ρ increases the zero-attracting effect (driving the estimate towards zero). Therefore, we test the zero-attracting effect for different levels of sparsity. The SNR is 3 dB in all cases.

(2) The effect of sparsity on the performance. We change the number of nonzero coefficients K to change the level of sparsity. The SNR is 3 dB in all cases.

(3) The effect of signal-to-noise ratio on the performance. We change the SNR of the observation noise. The underlying system has 3 nonzero coefficients in all cases.

The estimation performance is evaluated based on the ℓ_2-norm error with the true parameters, i.e., $||\widehat{\boldsymbol{h}}_{\rm FIN} - \boldsymbol{h}||_2$, where $\widehat{\boldsymbol{h}}_{\rm FIN}$ is the estimated $\widehat{\boldsymbol{h}}$ at the final iteration. The results are averaged over 1000 random trials, so as to obtain an average performance. The results are collected in Table 1. More details on the experiments and the algorithms used are in Appendix B.

Discussion on the results. The comparisons for different levels of regularization and sparsity show that the proposed method strikes a good trade-off between attracting the estimates towards zero (useful in sparse environments, e.g., $K = 3$) and providing estimates close to the true $\boldsymbol{h}$. In fact, as the level of sparsity decreases (e.g., $K = 5$ or $K = 7$), attracting the estimate towards zero may increase the estimation error, since many elements of $\boldsymbol{h}$ are different than zero. The proposed ℓ_1^2-RLS outperforms the other methods in most scenarios. The comparisons for different levels of SNR demonstrate that the proposed ℓ_1^2-RLS behaves better in all noisy situations. It is only when the SNR reaches 10 that ZA-RLS behaves as well as the proposed method. This good performance can be explained with the fact that the proposed cost (2) allows a minimization approach analogous to recursive least squares with Tikhonov regularization, which behaves well in sparse and non-sparse scenarios. Thus, an interesting future study is to dynamically change the size of the vector to be estimated, i.e., to reduce or increase it online depending on an estimated degree of sparsity.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62073074), Key Intergovernmental Special Fund of National Key Research and Development Program (Grant No. 2021YFE0198700), and Research Fund for International Scientists (Grant No. 62150610499).

Supporting information Appendixes A and B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Haykin S. Adaptive Filter Theory. 5th ed. Upper Saddle River: Pearson Prentice Hall, 2013

2 Tang Q, Li B Q, Chai Y, et al. Improved sparse representation based on local preserving projection for the fault diagnosis of multivariable system. Sci China Inf Sci, 2021, 64: 129204

3 Moustakis N, Zhou B, Quang T L, et al. Fault detection and identification for a class of continuous piecewise affine systems with unknown subsystems and partitions. Int J Adapt Control Signal Process, 2018, 32: 980–993

4 Chen Y, Gu Y, Hero A O. Sparse LMS for system identification. In: Proceedings of IEEE International Conference on Acoustics, Speech & Signal Processing, 2019. 3125–3128

5 Lim J, Lee K, Lee S. A modified recursive regularization factor calculation for sparse RLS algorithm with l_1-norm. Mathematics, 2021, 9: 1580

6 Eksioglu E M. Sparsity regularised recursive least squares adaptive filtering. IET Signal Process, 2011, 5: 480–487

7 Hong X, Gao J, Chen S. Zero attracting recursive least squares algorithms. IEEE Trans Veh Technol, 2016, 66: 213–221

8 Li Z F, Li D, Zhang J Q. A new penalized recursive least squares method with a variable regularization factor for adaptive sparse filtering. IEEE Access, 2018, 6: 31828–31839

9 Zou H, Zhang H H. On the adaptive elastic-net with a diverging number of parameters. Ann Stat, 2009, 37: 1733–1751