Convex regularized variable-forgetting-factor recursive least squares algorithm for sparse system identification

Xin Yang*, Yuhong Wang, Site Mo and Gen Wang

A convex regularized variable-forgetting-factor recursive least squares algorithm (CR-VFFRLS) is proposed for sparse system identification, in which the variable-forgetting-factor is deduced by minimizing the convex regularized cost function via the gradient descent method. It overcomes the drawback that the fast-tracking ability with the high steady-state error or the low steady-state error with slow tracking ability, which is ineluctable in the fixed forgetting-factor RLS algorithm. Simulation results demonstrate the superiorities of the proposed algorithm to the CR-RLS and VFFRLS algorithm.

Introduction: In recent years, sparse system identification has gained considerable attention due to the real-world applications (e.g. digital TV transmission channel and the echo paths). At the same time, sparse-aware adaptive algorithms have also been successfully developed, and achieved better performance than their traditional versions. The recursive least squares (RLS) algorithm is given much attention due to its convergence performance is superior to the least mean squares (LMS) algorithm in steady-state misalignment and convergence rate, at the expense of high computational complexity [1]. To further develop the convergence properties of the RLS algorithm in the sparse system identification scenario, the convex regularized RLS (CR-RLS) is suggested in [2], which adopts a general convex function of the system impulse response estimate in its cost function. However, the performance of RLS including tracking ability, steady-state misalignment, convergence speed and algorithm stability are governed by the forgetting-factor (FF). On the one hand, a large FF can make the algorithm achieve low steady-state misalignment and high algorithm stability. On the other hand, a small FF can make the algorithm obtain fast tracking ability and high convergence rate. Hence, a fixed FF must face a trade-off between the two hands. To solve the issue, a new control mechanism of the FF is presented to minimize mean squares error (MSE) by using the gradient of the MSE [3], but it comes up with numerical instability. Moreover, the robust variable-forgetting-factor RLS (VFF-RLS) is presented [4] by using the relationship between prior error and posterior error to adjust FF. The VFF-RLS provides a variable-forgetting-factor to meet the requirement of the two hands. However, it would face the problem that the estimation of the variance of the error signal and the observation noise are inaccurate, which degrades the performance of the VFF-RLS algorithm.

Inspired by the scheme of variable-forgetting-factor, we develop a convex regularized variable-forgetting-factor recursive least squares (CR-VFFRLS) algorithm for sparse system identification. The derivation of VFF is to minimize the cost function of the CR-RLS algorithm by using the gradient descent method. The simulation results confirm the efficiency of the CR-VFFRLS algorithm.

Conventional CR-RLS algorithm: Considering an unknown system whose input-output relationship is as follows

$$d(n) = \mathbf{w}_0^{\mathrm{T}}\mathbf{x}(n) + v(n) \tag{1}$$

where $\mathbf{w}_0$ denotes the unknown system parameter vector, $d(n)$ is the desired signal, $\mathbf{x}(n) = [\mathbf{x}(n)\ \mathbf{x}(n-1)\ \mathbf{x}(n-2),\ldots,\mathbf{x}(n-M+1)]^{\mathrm{T}}$ represents the input signal of n-time, $v(n)$ is the observation noise at time n.

The CR-RLS algorithm is defined by the following relations

$$e(n) = d(n) - \mathbf{w}^T(n-1)\mathbf{x}(n) \tag{2}$$

$$\mathbf{K}(n) = \frac{\mathbf{P}(n-1)\mathbf{x}(n)}{\lambda + \mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{x}(n)} \tag{3}$$

$$\mathbf{P}(n) = \frac{1}{\lambda}[\mathbf{P}(n\text{-}1) - \mathbf{K}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n\text{-}1)] \tag{4}$$

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{K}(n)e(n) - \gamma(1-\lambda)\mathbf{P}(n)\frac{df(\mathbf{w})}{d\mathbf{w}(n-1)} \tag{5}$$

$e(n)$ denotes the prior error parameter of n-time, $\mathbf{K}(n)$ is the Kalman gain vector. $\mathbf{P}(n)$ represents the autocorrelation matrix of the input vector, $\lambda \in (0,1)$ is the fixed forgetting factor. $\mathbf{w}(n) = [\mathbf{w}_1(n), \mathbf{w}_2(n), \mathbf{w}_3(n), \cdots, \mathbf{w}_M(n)]^{\mathrm{T}}$ is the estimation of the unknown parameter vector $\mathbf{w}_0$. γ is a regularization parameter. $f(\mathbf{w})$ is a general convex function.

Proposed CR-VFFRLS algorithm: When the unknown system is sparse, using the prior knowledge of system sparsity can obviously improve the convergence performance of the algorithm. Hence, the variable-forgetting-factor RLS algorithm for sparse system identification is proposed in the section. A constrained convex regularized cost function is defined as

$$\begin{aligned} &\underset{\lambda(n-1)}{\text{minimize}} \sum_{i=1}^{n} \lambda^{n-i}(n-1)[d(i) - \mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)]^2 + \gamma f(\mathbf{h}) \\ &\text{subject to } \lambda^2(n-1) < \xi^2 \end{aligned} \tag{6}$$

γ is a constant representing a compromise between the estimation error and the sparsity of the unknown system parameter. $\mathbf{h}$ is the estimation of the unknown parameter vector $\mathbf{w}_0$.

Converting (6) into an unconstrained optimization problem by using the method of Lagrange multipliers, we can obtain

$$\begin{aligned} J(\lambda(n-1)) = &\sum_{i=1}^{n} \lambda^{n-i}(n-1)[d(i) - \mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)]^2 \\ &+ \gamma f(\mathbf{h}) - \alpha(\lambda^2(n-1) - \xi^2) \end{aligned} \tag{7}$$

The partial derivatives for $\lambda(n-1)$ is expressed as

$$\begin{aligned} \frac{\partial J(\lambda(n-1))}{\partial\lambda(n-1)} = &\sum_{i=1}^{n}\{(\mathrm{n-i})\lambda^{n-i-1}(n-1)[d(i) - \mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)]^2 \\ &+ 2\lambda^{n-i}(n-1)[d(i) - \mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)](-\mathbf{x}(i))\frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)}\} \\ &+\gamma\frac{\partial f(\mathbf{h})}{\partial\mathbf{h}(n)}\frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)} - 2\alpha\lambda(n-1) \end{aligned} \tag{8}$$

where

$$\begin{aligned} \frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)} = &e(n)\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)} + \gamma\frac{df(\mathbf{h})}{d\mathbf{h}(n-1)}\mathbf{P}(n) \\ &-\gamma(1-\lambda(n-1))\frac{df(\mathbf{h})}{d\mathbf{h}(n-1)}\frac{\partial\mathbf{P}(n)}{\partial\lambda(n-1)} \end{aligned} \tag{9}$$

There

$$\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)} = -\frac{\mathbf{P}(n-1)\mathbf{x}(n)}{(\lambda(n-1) + \mathbf{x}^T(n)\mathbf{P}(n-1)\mathbf{x}(n))^2} \tag{10}$$

$$\begin{aligned} \frac{\partial\mathbf{P}(n)}{\partial\lambda(n-1)} = &-\frac{1}{\lambda^2(n-1)}[\mathbf{P}(n-1) - \mathbf{K}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n-1)] \\ &+\frac{1}{\lambda(n-1)}\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n-1)\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)} \end{aligned} \tag{11}$$

Using the gradient descent method, the variable-forgetting-factor is updated as

$$\lambda(\mathrm{n}) = \lambda(n-1) - \beta\frac{\partial J(\lambda(n-1))}{\partial\lambda(n-1)} \tag{12}$$

To guarantee the stability of the algorithm, the forgetting-factor should be bounded, in other words, it should be chosen in the range $[\lambda_{\min}, \lambda_{\max}]$. $\lambda_{\min}$ could provide a fast-tracking ability and a fast convergence rate, and $\lambda_{\max}$ could provide a low steady-state misalignment and high algorithm stability.

Considering the simplicity of the algorithm, a convex relation with L_1-norm constraint is used.

$$f(\mathbf{h}) = \|\mathbf{h}\|_1 \tag{13}$$

The corresponding derivative is expressed as

$$\mathrm{G}(\mathbf{h}) = \frac{df(\mathbf{h})}{d\mathbf{h}} = \mathrm{sgn}(\mathbf{h}) \tag{14}$$

$\mathrm{sgn}(\cdot)$ denotes the sign function.

The VFF-CRRLS algorithm is summarized in Table 1.

Table 1 Summary of VFF-CRRLS algorithm

Initialization:

$\mathbf{P}(0)=\sqrt{\delta}\mathbf{I}$, δ is a small positive constant;

$\mathbf{h}(0)=\mathbf{0}$, γ , $\mathbf{x}(n)$, $d(n)$, $\lambda_{\min}$, $\lambda_{\max}$, β

1. for n =1, 2, 3,…, N

2. $e(n)=d(n)-\mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(n)$

3. $$\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)}=-\frac{\mathbf{P}(n-1)\mathbf{x}(n)}{(\lambda(n-1)+\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n-1)\mathbf{x}(n))^2}$$

4. $$\frac{\partial\mathbf{P}(n)}{\partial\lambda(n-1)}=-\frac{1}{\lambda^2(n-1)}[\mathbf{P}(n\text{-}1)\text{-}\mathbf{K}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n\text{-}1)]+\frac{1}{\lambda(n-1)}\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n\text{-}1)\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)}$$

5. $$\frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)}=e(n)\frac{\partial\mathbf{K}(n)}{\partial\lambda(n-1)}+\gamma\,\mathrm{G}(\mathbf{h}(\mathrm{n}-1))\mathbf{P}(n)-\gamma(1-\lambda(n-1))\,\mathrm{G}(\mathbf{h}(\mathrm{n}-1))\frac{\partial\mathbf{P}(n)}{\partial\lambda(n-1)}$$

6. $$\psi(n)=\sum_{i=1}^{n}\{(n-\mathrm{i})\lambda^{n-i}(n-1)[d(i)-\mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)]^2+2\lambda^{n-i}(n-1)[d(i)-\mathbf{h}^{\mathrm{T}}(n)\mathbf{x}(i)](-\mathbf{x}(i))\frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)}\}+\gamma\,\mathrm{G}(\mathbf{w}(\mathrm{n}-1))\frac{\partial\mathbf{h}(n)}{\partial\lambda(n-1)}-2\alpha_2\lambda(n-1)$$

7. $$\mathbf{K}(n)=\frac{\mathbf{P}(n-1)\mathbf{x}(n)}{\lambda(n-1)+\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n-1)\mathbf{x}(n)}$$

8. $\mathbf{P}(n)=\lambda^{-1}(n-1)[\mathbf{P}(n-1)-\mathbf{K}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{P}(n-1)]$

9. $\lambda(\mathrm{n})=\lambda(n-1)-\beta\psi(n)$

10. $\mathbf{h}(n)=\mathbf{h}(n-1)+\mathbf{K}(n)e(n)-\gamma(1-\lambda(n))\mathbf{P}(n)\mathrm{G}(\mathbf{h}(\mathrm{n}-1))$

End for

Simulation results: To verify the convergence performance of the proposed VFF-CRRLS algorithm, computer simulation is tested for sparse system identification. The unknown sparse system parameter is defined as $\mathbf{w}_0=[1,\underbrace{0,0,\ldots}_{M/4-1},1,\underbrace{0,0,\ldots}_{M/4-1},1,\underbrace{0,0,\ldots}_{M/4-1},1,\underbrace{0,0,\ldots}_{M/4-1}]^{\mathrm{T}}$. The unknown system parameter suddenly changes to $-\mathbf{w}_0$ at every 1/4 of the sum samples. The input signal $x(n)$ is a zero-mean white Gaussian signal with variance $\sigma_x^2=1$. The background noise $v(n)$ is a Gaussian signal with zero-mean and variance $\sigma_v^2=1$, in which the signal-to-noise ratio (SNR) is 20dB. The normalized mean-square-deviation (NMSD) is used to evaluate the performance of the algorithms, which is defined as

$$\mathrm{NMSD}=10\log_{10}[\|\mathbf{h}(n)-\mathbf{h}_0\|_2^2/\|\mathbf{h}_0\|_2^2] \quad (15)$$

We set the parameters as: $\lambda=0.995$ for CR-RLS, $K_\alpha=3$, the forgetting-factor range $[0.001,0.999999]$, $\gamma=2$ for VFF-RLS, $\gamma=0.2$, $\beta=4\times10^{-4}$, the forgetting-factor range $[0.8,0.999999]$.

Fig.2 shows the NMSD performance of the CR-RLS, VFF-RLS and the proposed CR-VFFRLS algorithm. As can be seen, the VFF-RLS algorithm has a better tracking-ability than the CR-RLS. The proposed CR-VFFRLS algorithm has the best performance in terms of tracking ability, steady-state error. The forgetting-factor curves of the two VFF algorithms are shown in Fig.1, obviously, the FF of the proposed CR-VFFRLS stays at the minimum for more iterations when the unknown system changes.

Conclusion: In the paper, a convex regularized variable-forgetting-factor recursive least squares (CR-VFFRLS) algorithm has been proposed for sparse system identification. The derivation of VFF is based on minimizing the convex regularized cost function. To illustrate the properties of the CR-VFFRLS algorithm, the comparison between VFF-RLS and the proposed CR-VFFRLS algorithm in FF is studied. Simulation results show the advantages of the CR-VFFRLS algorithm.

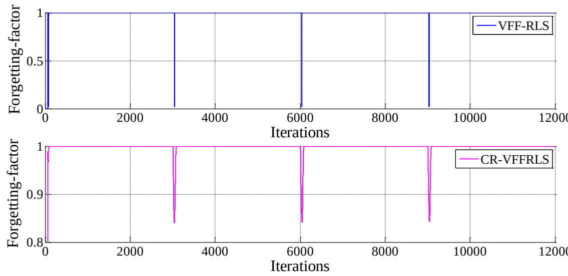

Fig.1 *Comparison between VFF-RLS and CR-VFFRLS in FF.*

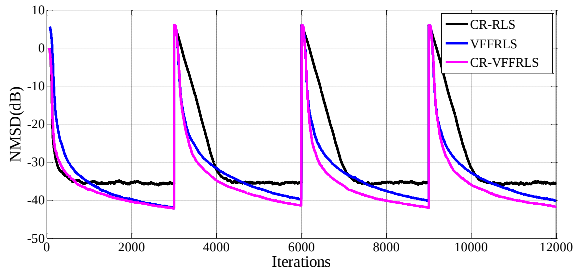

Fig.2 *The NMSD learning curves of the CR-RLS, VFF-RLS and CR-VFFRLS algorithm.*

Acknowledgments: Our work was partially supported by National Science Foundation of P.R. China (Grant: 61871461).

Xin Yang and Gen Wang *(State Grid Deyang Power Supply Company, Deyang, 618400, People's Republic of China)*
Yuhong Wang and Site Mo (*School of electrical engineering, Sichuan University, Chengdu, 610065, People's Republic of China*).

* Corresponding author
Email: 2569362745@qq.com.

Reference

1 Haykin, S., Adaptive Filter Theory. Prentice Hall, Upper Saddle River, NJ (2002)

2 Eksioglu, E. M. and Tanc, A. K.: 'RLS Algorithm With Convex Regularization', in *IEEE Signal Processing Letters*, 2011, **18** (8), pp. 470-473

3 Leung, S. H. and So, C. F.: 'Gradient-based variable forgetting factor RLS algorithm in time-varying environments', *IEEE Trans. Signal Process.*, 2005, **53 (**8), pp. 3141–3150

4 Paleologu, C., Benesty, J. and Ciochina, S.: 'A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification', *IEEE Signal Processing Letters*, 2008, **15**, pp. 597-600