Proc. ISPA, pages 107-112, Dubrovnik, Croatia, September 2011, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6046589

Noise Resilience Through Band-Limitation in Signal Regression Analysis

Rade Kutil
University of Salzburg, Department of Computer Sciences
Jakob Haringer-Str. 2, 5020 Salzburg, Austria
Email: rkutil@cosy.sbg.ac.at

***Abstract*—Linear regression is used in signal analysis when other methods like artificial neural networks or support vector machines either lack the ability to represent the result in form of a signal or cannot be applied to continuous target values. However, signal noise may lead to unstable noisy solutions with bad performance on non-trained data, especially for underdetermined systems. This work develops a method to add statistical virtual noise with special properties such as band-limitation to the signals in order to reduce these properties in the solution signal. The results show stable solutions with significantly improved performance on non-trained data. The method is also tested on real EEG data.**

I. Introduction

Regression analysis seeks to calculate a target value for each instance of source data. The target values are given for a set of instances called the training set. Thus, the system learns which properties of the source data can be used to deduce the target value.

There are a lot of schemes to achieve this which all have certain advantages and disadvantages. For reasons of easy use and computational performance, linear methods are very popular. Here, the system optimizes the coefficients of a linear combination of the source data values in order to achieve minimal deviations from the target values.

In signal analysis, the source data consists of sampled signal data. A set of signals is used as the training set to learn optimal parameters. The signals can contain noise that is not related to the target values. Therefore, the analysis procedure should automatically exclude the noise in order not to produce unstable results on non-trained signal data. However, this is not always possible. Robust methods [1], [2] have been developed to cope with this problem.

Linear regression mostly minimizes the squared error (least squares). Robustness against noise can either be achieved by excluding statistical outliers as in the RANSAC algorithm [3], or by using other estimators such as in the least-median-of-squares method [4], or by using M-estimators [5]. All these methods concentrate on the outlier-problem. However, noise can affect all samples, so other methods might be more suitable for signal analysis.

Another important linear scheme is the support vector machine [6]. It does not minimize the squared errors but finds a sub-plane that best separates classes of source data instances. It turns out to yield good classification performance and is quite robust against noise. However, in under-determined situations, i.e. where that training set is smaller than the signal length, near-plane data has a major influence on the outcome and can make the result unstable. Moreover, it cannot be used on continuous target values since it is a classification scheme.

There are also non-linear schemes such as artificial neural networks [7], [8]. There are two problems with those. First, it is hard to get a grip on the learning space. This leads to suboptimal local minima and uncontrollable overfitting. Second, they lack the nice property of linear methods that the learned coefficients can be arranged as a time series and are, thus, interpretable as a signal.

The motivation of this work comes from the analysis of EEG signals [9] from experiments in neurosciences. Mostly, the task is the classification of those signals [10], [11] according to the experimental setup and subject responses. The resulting solution signals are interpreted in a physiological sense [12].

EEG signals are very noisy. Usually, a number of signal instances (single trials) are averaged to form the so called event related potential (ERP). When only signals with a certain target value are averaged, the corresponding signal properties add up, while noise does not. This is basically a crude form of a linear regression analysis. Therefore, attempts have been made to use proper linear regression.

To cope with the noise in EEG signals, conventional noise robust schemes seemed not to do the trick. Therefore, statistical signal properties such as the expected frequency bands of the useful signal components (θ-, α-, and γ-oscillations) have been exploited in [13] and finally [14], which use the methods described in this work.

II. Noise Resilience Through Artificial Statistical Noise

The big problem for linear regression on noisy data is that the result vector very often relies more on noise than on the underlying signal to approximate the target vector. This leads to result vectors that are not only noisy themselves and do not represent a reasonable aspect of the original data, they are also very unstable when tested on data they have not been trained with.

There can be several reasons for unusable noisy result vectors:

- The whole equation system might be underdetermined, i.e. the training set is smaller than the number of data

points in the signal. This leads to a certain degree of freedom which is immediately filled by noise.
- Even if the equation system is not underdetermined, the underlying (noiseless) signals might be, ironically, too uniform. This would lead to a noiseless equation system with low rank, again implying degrees of freedom that are filled by noise.
- If the signals are band-pass filtered to remove unwanted noise prior to performing regression analysis, the opposite is achieved for the result vector. This is because the product of the signals and the result vector must, out of principle, stay the same. So, whatever is suppressed in signal data is boosted in the result vector, which means noise.

To remove noise outside a certain frequency range $[f_0, f_1]$ from the solution vector, we might either filter the solution vector, which might have unexpected effects on the regression error, or use the following method. Contrary to filtering the source data, we virtually increase the data's noise outside the desired frequency range.

A. Real Signals

The known result of linear regression is that the least square solution of $\sum_t a(t)x(t) \approx b$, where a is a random signal and b a random target, can be found by solving the square linear system

$$\sum_t E(a(s)a(t))x(t) = E(a(s)b)\,. \tag{1}$$

Now we substitute the signal $a(t)$ by $\tilde{a}(t) = a(t)+n(t)$, where $n(t)$ is a special noise signal that will serve our needs, with $E(n(t)) = 0$. Thus, the correlation matrix becomes

$$E(\tilde{a}(s)\tilde{a}(t)) = E(a(s)a(t)) + E(n(s)n(t)) \tag{2}$$

which means that we just have to add the correlation matrix of the additional noise to our linear system.

In case of white noise, the correlation matrix has diagonal form and the diagonal elements contain the variance of the noise at a single point t. Therefore, by adding a certain positive value to the diagonal of the correlation matrix, a higher number of signals in the training set is simulated. The larger the added value, the more resilient to noise the result should be. However, the resulting solution $x(t)$ will also be more conservative, i.e. it will tend to produce flat signals around the average target value. This method has been used in [13]. However, experiments in [13] have shown that a significant improvement is only achieved for underdetermined systems. Moreover, an increase of the value added to the diagonal only slightly improves classification rates, if at all. Therefore, only a small value should be used, in order to make the correlation matrix have full rank, but not to decrease the power of the solution signal.

To avoid suppressing useful data in the same way as noise, we may model the added statistical noise so that it has properties contrary to that of the original signals, which is known in many applications. In our case, we want the noise to fill a frequency spectrum outside of the expected signal frequency band.

Thus, initially we model the noise by a single oscillation with frequency $f = 2\pi\omega$, assuming, for the moment, a sampling frequency of 1.

$$n_\omega(t) = X\cos\omega t + Y\sin\omega t\,, \tag{3}$$

where X and Y are two independent random variables with identical distributions and $E(X) = E(Y) = 0$. We get

$$E(n_\omega(s)n_\omega(t)) = \frac{1}{2}(E(X^2) + E(Y^2))\cos\omega(t-s)\,, \tag{4}$$

where we will abbreviate the constant factor $\frac{1}{2}(E(X^2) + E(Y^2))$ by A. Not enough, we want our noise to fill an entire band of frequencies $[f_0, f_1]$. Therefore, we model X and Y as dependent on ω but with constant A and integrate over the frequency range to get our final noise correlation

$$\int_{\omega_0}^{\omega_1} E(n_\omega(s)n_\omega(t)) = A(\omega_1 \operatorname{sinc}\omega_1(t-s) - \omega_0 \operatorname{sinc}\omega_0(t-s)) \tag{5}$$

Now, as we want to add noise that fills the frequencies outside of $[f_1, f_2]$, we add two noise bands $[0, f_1]$ and $[f_2, f_s/2]$, where f_s is the sampling frequency (here 1). Thus, what we have to add to our correlation matrix is

$$E(\tilde{a}(s)\tilde{a}(t)) = E(a(s)a(t)) + \begin{cases} A(\pi - \omega_2 + \omega_1) & s = t \\ A\frac{\sin\omega_1(t-s) - \sin\omega_2(t-s)}{t-s} & s \neq t\,. \end{cases} \tag{6}$$

The solution vector should then be mainly in the frequency range $[f_1, f_2]$, with other frequencies suppressed for high enough values of A.

B. Complex Signals

In case of complex signals, e.g. Gabor-transformed signals, we want to solve $\sum_t a(t)\overline{x(t)} \approx b$, where $a(t)$ is a complex random signal and b is the complex target, although b is real in our test cases. The corresponding least square solution is found by

$$\sum_t E(a(s)\overline{a(t)})x(t) = E(a(s)\bar{b})\,, \tag{7}$$

where we, again, substitute $\tilde{a}(t) = a(t) + n(t)$ for $a(t)$ to add statistical noise $n(t)$. For a certain frequency $2\pi\omega$, the complex noise signal n_ω can be modeled as

$$n_\omega(t) = Xe^{i\omega t}\,, \tag{8}$$

where X is a complex random variable with zero mean. Then

$$E(n_\omega(s)\overline{n_\omega(t)}) = E(|X|^2)e^{i\omega(s-t)}\,, \tag{9}$$

and in the same way

$$\int_{\omega_0}^{\omega_1} E(n_\omega(s)\overline{n_\omega(t)}) = \begin{cases} A(\omega_1 - \omega_0) & s = t \\ A\frac{e^{i\omega_1(s-t)} - e^{i\omega_0(s-t)}}{i(s-t)} & s \neq t\,, \end{cases} \tag{10}$$

which has to be added to the correlation matrix.

For complex signals, we can distinguish between positive and negative frequencies. We might want our solution signal to have only positive frequencies in the band $[f_1, f_2]$. Therefore, we add two noise bands $[-f_s, f_1]$ and $[f_2, f_s]$, where f_s is the sampling frequency. Thus, the complex correlation matrix becomes

$$E(\tilde{a}(s)\overline{\tilde{a}(t)}) = E(a(s)\overline{a(t)}) + \begin{cases} A(2\pi - \omega_2 + \omega_1) & s = t \\ A\frac{-e^{i\omega_2(s-t)} + e^{i\omega_1(s-t)}}{i(s-t)} & s \neq t\,. \end{cases} \tag{11}$$

For Gabor-transformed signals, there are several frequency bands, which should conform to a certain smoothness in time according to the frequency f of the band. Therefore, we build a frequency range, e.g. $[f-3, f+3]$, around this frequency and apply this scheme for each band. What remains is "vertical" noise in the frequency domain. Therefore, we apply the scheme again in the frequency dimension for each point in time.

III. Artificial Test-Set

In order to assess the benefits of the suggested noise suppression scheme in a controlled test environment, we synthesize a test signal $\tilde{a}(t)$ consisting of a sinusoid with amplitude 1 and randomized phase and white noise with maximum amplitude N.

$$\tilde{a}(t) = a(t) + N(\mathrm{rand}(1) - 0.5)\,, \tag{12}$$

where $\mathrm{rand}(r)$ generates a random number in the interval $[0, r]$, and

$$a(t) = \begin{cases} \sin(2\pi f t + \phi) & b = 1 \\ \sin(2\pi f t + \pi + \phi) & b = 0\,, \end{cases} \tag{13}$$

where $\phi = \mathrm{rand}(0.5)$ is the random phase of the test signals, which is constant for a single signal but changes for each signal. The signals are sampled at 500 samples at sampling rate 1. The frequency is chosen to have 10 periods within the signal length, i.e. $f = \frac{10}{500}$.

The linear regression system is trained with 1000 signals, 500 each for $b = 1$ and $b = -1$. The frequency band for noise suppression is chosen as $[\frac{7.5}{500}, \frac{12.5}{500}]$. Fig. 1 shows an example solution signal for noise strength $N = 20$. One can see that the solution signal is extremely noisy in the unmodified case. The noise suppression scheme manages to pull a smooth signal out of noisy signal data without reducing the solution signal amplitude. This improves classification performance, as seen below. Furthermore, the signal is much more expressive for analysis purposes, since frequency and phase can easily derived from it.

Note that the simple stabilization method of adding a constant value to the diagonal of the correlation matrix is not applied here because the system is overdetermined and, thus, the method does not improve the solution signal.

The solution signal is then tested on 1000 test signals of the same form as the training signals, but with different

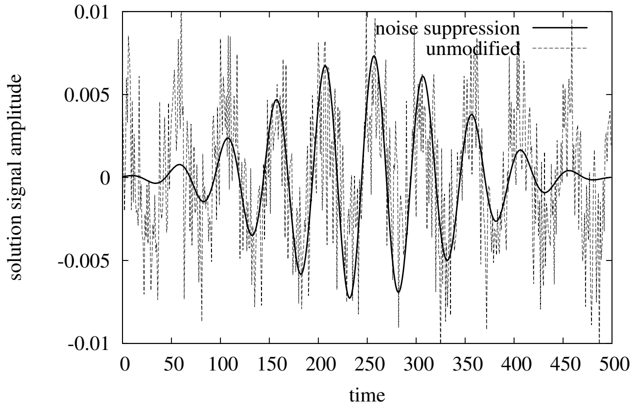

Fig. 1. Solution signals for artificial test set. The noise strength in comparison to signal strength is $N = 20$.

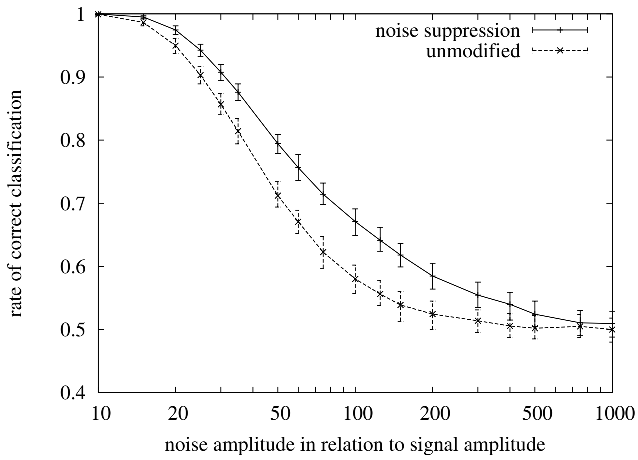

Fig. 2. Results of classification test for over-determined artificial test set depending on noise strength.

random values, again 500 each for $b = 1$ and $b = -1$. If the sign of $\sum_t \tilde{a}(t)x(t)$ is the same as that of b, then the signal counted as correctly classified. The rate of correctly classified out of all test signals is computed and displayed in Fig. 2. Actually, the value displayed in Fig. 2 is an average computed over 101 of such tests for each noise strength to eliminate random deviations. Additionally, the lowest and highest decile is plotted as error bars to show how large the deviations from the average value actually are.

One can see that the noise suppression scheme is able to increase the correct classification rate by up to 10 percent points, depending on the noise strength. For noise strength $N = 10$ and below, both noise suppressed as well as unmodified solution signals detect almost all signals correctly. For $N = 500$ and above, the classification rate of both schemes is understandably around 50%, i.e. their output is that of basically random guesses. For medium noise strength, the improvement of our noise suppression scheme is greatest at about $N = 100$. This means that, if the noise is about 100 times as strong as the original signal, the classification rate

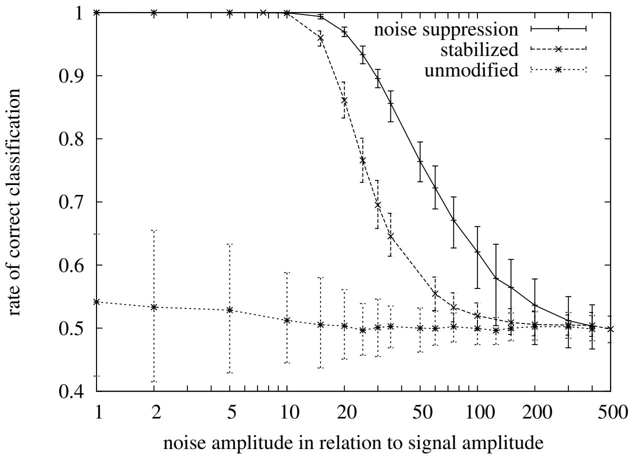

Fig. 3. Results of classification test for under-determined artificial test set depending on noise strength.

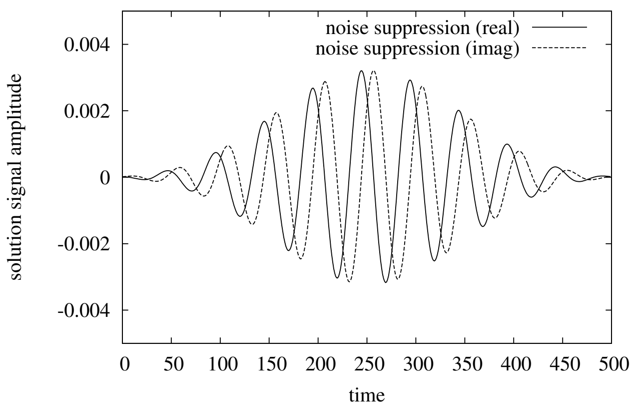

Fig. 4. Complex solution signal for artificial test set. The noise strength in comparison to signal strength is $N = 20$. The unmodified solution signal is too noisy to be shown in this plot.

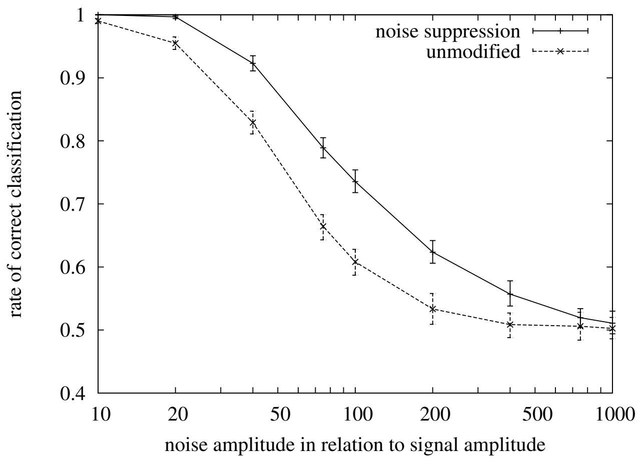

Fig. 5. Results of complex classification test for over-determined artificial test set depending on noise strength.

can be increased from 57% to 67%.

So far, the solution has been over-determined by the test data, i.e. the equation system has been built of $2n$ signals, so there are twice as many equations than variables (points in $x(t)$), and the least squares system gives a unique solution. But what if we have a very small training set, say 20 signals, 10 for $b = 1$ and 10 for $b = 0$?

Fig. 3 shows the results for this case. The unmodified system basically produces arbitrary noise signals that are not capable of classification rates that significantly exceed the random guess level of 0.5 unless the noise in the training and test signals is very low.

Therefore, we apply a soft statistical white noise to stabilize the least squares equation system, i.e. we add a small value, 0.001, to the diagonal of the correlation matrix. In this way, the solution is unique and we get reasonable classification rates up to noise strength $N = 50$. Greater values do not produce better classification rates.

However, the band-limiting noise suppression is able to improve these rates by up to over 20 percent points and gives significant results up to noise strength $N = 200$. This shows that it is even more important to exploit knowledge about signal characteristics, such as limited frequency bands, when the training set is small, in order to improve results in regression analysis.

Now for complex signals. We synthesize the complex test signal $\tilde{a}(t)$ in a similar way:

$$\tilde{a}(t) = a(t) + N(\mathrm{rand}(1) - 0.5) + iN(\mathrm{rand}(1) - 0.5) \quad (14)$$

where

$$a(t) = \begin{cases} e^{i2\pi ft+\phi} & b = 1 \\ e^{i2\pi ft+\pi+\phi} & b = 0\,, \end{cases} \quad (15)$$

and $\phi = \mathrm{rand}(0.5)$ as before. All other parameters also remain the same.

Fig. 4 shows a sample solution signal for noise strength $N = 20$. While the solution of the unmodified system almost entirely consists of random noise, noise suppression yields a smooth signal with the correct frequency and phase.

Again, we test the solution signal of a training set of size $2n$ for a range of noise strengths and calculate average classification rates, where a signal $a(t)$ counts as correctly classified to b by $x(t)$ if the sign of the real part of $\sum_t a(t)\overline{x(t)}$ is the same as that of b. Fig. 5 shows the results. Noise suppression achieves an improvement of classification rates of up to almost 15 percent points, which is even better than for real signals.

Finally, in the under-determined case of only 10+10 training signals the situation is similar to the real case. Fig. 6 shows the results. The unmodified system produces complex random noise and the classification rates remain at the random guess level. Adding a real value of 0.001 to the diagonal of the correlation matrix, representing white noise, stabilizes also the complex system. However, the band-limited noise suppression is able to improve the classification rates by up to 25 percent points, which is again better than in the real case.

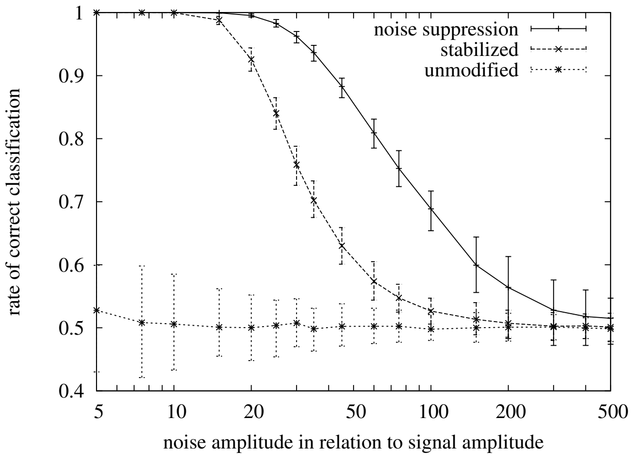

Fig. 6. Results of complex classification test for under-determined artificial test set depending on noise strength.

IV. Application to EEG Signals

The described methods have been tested [14] to reproduce findings from [15] in an experiment in neuroscience. The given experiment is a cued visual attention task. Subjects had to fixate the center of a computer monitor while an arrow indicated the appearance of a target stimulus for about 34 ms. After a duration ranging between 600 and 800 ms a target stimulus appeared for 50 ms on the left or on the right monitor side (the target was either a large or a small bar). The target had not to appear mandatory on the indicated monitor side – in 75% of the trials the arrow indicated the correct side where the target appeared, in 25% it indicated the wrong monitor side. The goal for the proband was to press a certain button for a small bar (target) and another button for a large bar.

EEG data has been recorded from several electrodes. In our case, electrode positions O1 and O2 are analyzed because processing of the task is expected there. See Fig. 7. Only valid trials are used, i.e. those where the target appears at the indicated side. If the target appears on the left side, it is processed on the right side an vice versa. The goal of the analysis is to determine what the difference between processing and non-processing is, i.e. what the signal properties indicate a contralateral as opposed to an ipsilateral situation.

To do so, the EEG signals of the 200 milliseconds after the stimulus from both electrodes are concatenated to the signal $a(t)$ and those with a left stimulus are associated with a positive target b, whereas right stimuli get a negative b. About 600 such signals (single trials) have been extracted for each subject, 300 with a left and 300 with a right target. At a sampling frequency of 250Hz, the signal length is 50 samples.

Fig. 8 shows the solution signal of the unmodified system for a single subject. The left part of the figure shows the O1 electrode of the signal, the right part shows the O2 electrode. The signal is very noisy and does not represent any neurological findings.

Fig. 9, however, shows the solution with noise suppression outside of the frequency band $[1\text{Hz}, 20\text{Hz}]$. The result clearly

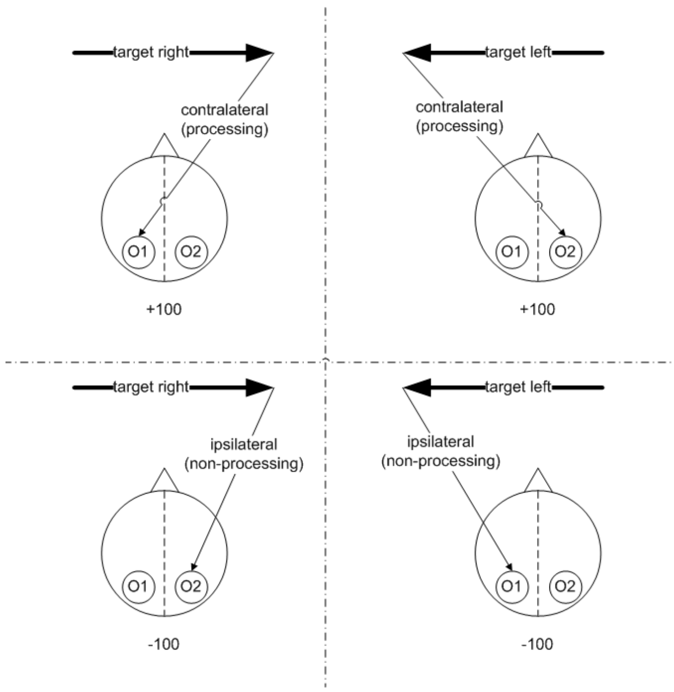

Fig. 7. Visual stimuli are processed contralaterally

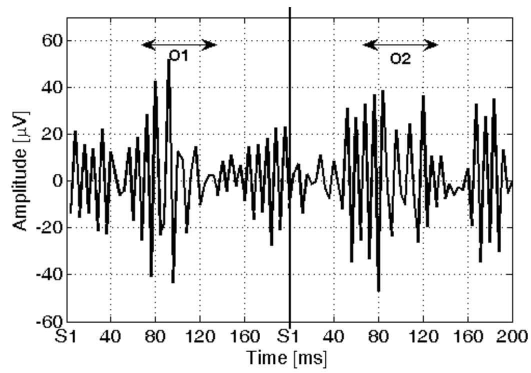

Fig. 8. Solution signal without modification of the correlation matrix

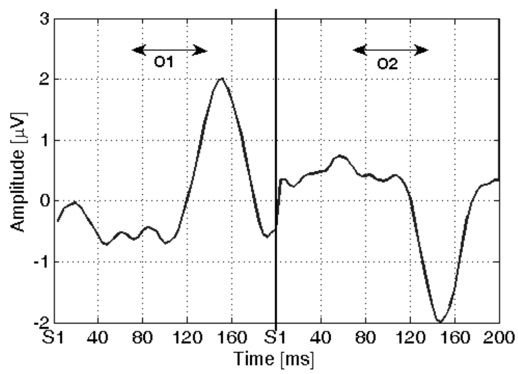

Fig. 9. Solution signal with noise suppression

TABLE I
EEG CLASSIFICATION RATES

Method	min	avg	max
unmodified	0.451	0.626	0.806
noise suppression	0.509	0.710	0.866

shows that the position of processing can be determined at 150ms after the stimulus, where O1 has a positive peak and O2 a negative peak for a left stimulus, which corresponds to the findings in [15].

To evaluate the classification rate on this data, a 1-out-of-n test is conducted, i.e. the system is trained with all but one single trials and then tested against the remaining one, which is repeated for each single trial and the rate of correctly classified single trials is calculated. Thus, we get a classification rate for each of the 22 subjects. Table I shows the minimum, average and maximum rates of all subjects. Like with artificial data, the performance could be improved by about 10 percentage points.

Note that the simple stabilization method of adding a constant value to the diagonal of the correlation matrix does not significantly improve the classification rates over the unmodified system because the system is overdetermined. Note also that the noise present in the data cannot be varied like in Fig. 3 because the noise comes from a real world measurement.

Several other analyses have been done on the test data, including Gabor-transformed signals as input to complex regression analysis over all subjects, where each Gabor coefficient was normalized in order to ignore inter-subject signal power differences. The goal was to prove a laterality-dependent latency of the EEG-signal, which should manifest itself in the signal phase. Although classification rates are low, since subject independent analysis is hard, the noise suppression scheme manages to improve the rates from 0.519 to 0.569.

V. Conclusion

It has been shown that it is possible to remove certain signal components from the solution signals of linear regression by adding virtual statistical noise to the source data. This works because whatever is emphasized in training data must be suppressed in the solution signal, since the product of the two must remain the same.

The way to do this is to add the correlation matrix of the virtual noise to the correlation matrix of the training data, which is used in the least squares solver of the linear regression scheme. So, in order to get a solution signal within a certain frequency range, noise outside this range has to be added. The correlation matrices of such noise signals have been found in closed form for real as well as complex signals.

Tests on an artificial test set in a classification problem show that the new method can increase the correct classification rate by 10 percentage points or more, depending on whether the system is under- or over-determined. Tests on real EEG data from an experiment in neuroscience confirm these results.

Future work might be to allow for transition bands in the virtual noise to avoid the dampening of the solution signals at the signal borders. Also, the frequency bands that best improve the classification rates might be found automatically, resulting in an unsupervised learning scheme.

Acknowledgment

The author would like to thank Michael Radlingmaier for implementing the framework [14] in which the proposed scheme has been tested, and the neuroscience working group of the Department of Psychology at the University of Salzburg for providing the experimental EEG data of [15].

References

[1] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust regression methods for computer vision: A review," *International Journal of Computer Vision*, vol. 6, pp. 59–70, 1991.

[2] M. C. Ortiz, L. A. Sarabia, and A. Herrero, "Robust regression techniques: A useful alternative for the detection of outlier data in chemical analysis," *Talanta*, vol. 70, no. 3, pp. 499–512, 2006.

[3] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, June 1981.

[4] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, Dec. 1984.

[5] J.-H. Chen, C.-S. Chen, and Y.-S. Chen, "Fast algorithm for robust template matching with M-estimators," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 230–243, Jan. 2003.

[6] V. Vapnik, *Estimation of dependences based on empirical data*. New York: Springer, 2006.

[7] I. A. Basheer and M. Hajmeer, "Artificial neuronal networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, Dec. 2000.

[8] C.-C. Lee, Y.-C. Chiang, C.-Y. Shih, and C.-L. Tsai, "Noisy time series prediction using m-estimator based robust radial basis function neural networks with growing and pruning techniques," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4717–4724, 2009.

[9] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 457–463, Dec. 2000.

[10] N. S. Dias, M. Kamrunnahar, P. M. Mendes, S. J. Schiff, and J. H. Correia, "Comparison of EEG pattern classification methods for brain-computer interfaces," *Engineering in Medicine and Biology Society*, pp. 2540–2543, 2007.

[11] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.

[12] P. Sauseng and W. Klimesch, "What does phase information of oscillatory brain activity tell us about cognitive processes?" *Neuroscience and Biobehavioral Reviews*, vol. 32, pp. 1001–1013, 2008.

[13] C. Wunderl, "Linear classification of EEG signals," Master's thesis, Paris-Lodron-University of Salzburg, Austria, Nov. 2005.

[14] M. Radlingmaier, "EEG signal feature extraction," Master's thesis, Paris-Lodron-University of Salzburg, Austria, Jan. 2011.

[15] R. Freunberger, Y. Höller, B. Griesmayr, W. Gruber, P. Sauseng, and W. Klimesch, "Functional similarities between the P1 component and alpha oscillations," *European Journal of Neuroscience*, vol. 27, pp. 2330–2340, 2008.