

The Use of the SPSA Method in ECG Analysis

László Gerencsér*, György Kozmann, Zsuzsanna Vágó, and Kristóf Haraszti

Abstract—The classification, monitoring, and compression of electrocardiogram (ECG) signals recorded of a single patient over a relatively long period of time is considered. The particular application we have in mind is high-resolution ECG analysis, such as late potential analysis, morphology changes in QRS during arrhythmias, T-wave alternants, or the study of drug effects on ventricular activation. We propose to apply a modification of a classical method of cluster analysis or vector quantization. The novelty of our approach is that we use a new distortion measure to quantify the distance of two ECG cycles, and the class-distortion measure is defined using a min-max criterion. The new class-distortion-measure is much more sensitive to outliers than the usual distortion measures using average-distance. The price of this practical advantage is that computational complexity is significantly increased. The resulting nonsmooth optimization problem is solved by an adapted version of the simultaneous perturbation stochastic approximation (SPSA) method of [1]. The main idea is to generate a smooth approximation by a randomization procedure. The viability of the method is demonstrated on both simulated and real data. An experimental comparison with the widely used correlation method is given on real data.

Index Terms—Classification, ECG, data-compression, min-max problems, monitoring, nonsmooth optimization, randomization, stochastic approximation, vector quantization.

I. INTRODUCTION

THE long-term measurement and real-time evaluation of biophysical signals is an essential task in several areas of medicine. In some applications, the characterization and classification of the signals is the basic goal, while in other cases the early detection of the significant changes is required. Furthermore, the need for compressing and archiving of data is an increasing demand both in the domain of health care and research.

Without confining the general applicability of our mathematical approach discussed in this paper, we consider primarily the problem of monitoring, compressing and archiving characteristic electrocardiogram (ECG) wave patterns obtained for a single patient over a relatively long period of time. Monitoring

the condition of a patient and keeping record of the patient's history in a compressed form is required in a number of situations, such as ambulatory ECG acquiring and processing heart-signals 24 hours a day (cf. [2]), exercise ECG, ECG taken during surgery or intensive care, or catheterization. ECG signals vary over time (beat-to-beat, day-to-day) even if the condition of the patient is stable, due to biological effects (positional changes due to breathing, variation in pulse rate etc.) and also due to technical reasons (e.g., noise of the measuring equipment). In most applications, we are interested in changes beyond the normal physiological variability. In this paper, we develop a method using a refined classification procedure that takes into account physiological variability. This work has been motivated by our interest in elaborating improved signal processing procedures to enhance the resolution of sophisticated ECG measurements, such as surface His-bundle ECG or late potential measurements. There are evidences that the tiny details hardly recognizable in surface ECG signals (or the temporal variation of such signal components) carry clinically meaningful indicators of elevated risk for malignant arrhythmias [3], [2], [4]. Furthermore, theoretical studies has proved, that based on an accurate detection of ventricular breakthrough points, the ill-posed problem of the inverse problem of electrocardiology can be transformed into a well-posed problem [5].

The main step in solving these problems is to generate a set of prototyped signals that represent a large set of real signals measured during the monitoring period. The generation of prototyped signals is a key mathematical challenge. Classical methods of cluster-analysis or vector quantization (cf. [6]–[8]) are not applicable for two main reasons. First, the right metric in the signal space is not Euclidean, in fact we are working in a constrained signal space. Second, averaging would mask valuable information contained in a relatively small percent of data.

In our procedure, the number of classes is not given *a priori*, instead, we define classes by setting an upper bound for the class distortion measure. We propose that prototypes or centers are defined so that the maximum of an appropriate distance function is minimized. This is a nonsmooth optimization problem the solution of which is not trivial for large data sets (cf. [9]). We propose a solution of this problem by using an adapted version of the so-called simultaneous perturbation stochastic approximation (SPSA) method, given in [1]. The convergence properties of this randomization method are well-understood and empirical evidence suggest that SPSA is applicable to solve difficult optimization problems. A main technical advance is the extension of the SPSA method to nonsmooth problems to get a suboptimal solution, which is certainly satisfactory for the present application. The viability of the proposed procedure will be demonstrated by numerical experiments for the solution of the min-max problem on both simulated and real data.

Manuscript received March 13, 2000. This work was supported by the National Research Foundation of Hungary (OTKA) under Grant T 032932, Grant T 033085, Grant T 030747, and Grant NKFP OM-2/052 and in part by a Bolyai János Research Fellowship. Asterisk indicates corresponding author.

*L. Gerencsér is with the Computer and Automation Institute of the Hungarian Academy of Sciences, Kende u. 13-17, H-1111 Budapest, Hungary (e-mail: gerencser@sztki.hu).

G. Kozmann and K. Haraszti are with the Department of Information Systems, University of Veszprém and the Research Institute for Technical Physics and Materials Science of the Hungarian Academy of Sciences, Budapest, Hungary.

Z. Vágó is with the Computer and Automation Institute of the Hungarian Academy of Sciences, Budapest, Hungary and the Pázmány Péter Catholic University, Budapest, Hungary.

Publisher Item Identifier 10.1109/TBME.2002.802007.

The integration of medical expertise into mathematical models can be done right at the start: by choosing an appropriate signal representation and by defining an appropriate metric between ECG signals. A variety of options will be explained in details in Section II. The approximate reconstruction of the full data is described in [10].

The ECG signal is an analog signal the value of which estimates the potential of the cardiac sources on the body surface. Several branches of experimental technologies have been elaborated during the long history of the electrocardiology, including body surface potential mapping as the most complete assessment of the heart-generated potential fields, as well as simplified methods, like the most popular standardized 12-lead ECG, the vector cardiography, VCG, high-resolution ECG, exercise ECG, etc. The most comprehensive review of all these techniques can be found in [11]. See, also, [12].

The usual duration of computer evaluated ECG records is 10 s. A digital ECG signal is obtained by sampling the analog signal at a rate of min. 500 Hz, with a usual resolution of 5 mV. Thus, the digital signal can be represented as a vector in a vector space of dimension min. $10 \times 500 = 5000/\text{lead}$. With 12 leads the digitized ECG signal would be represented as a vector of dimension min. $12 \times 500 = 60.000$. Computerized analysis of data of this dimension can not be performed without exploiting the structure in data and reducing dimensionality. Several data compression methodologies has been reported. The current clinical routine approaches keep one lead for arrhythmia information, such as RR distances, while the waveform information is derived only from one representative (or averaged) cycle of the majority cluster. For clinical classification purposes, parameter tables are filled out with approximately $12 \times 15 = 180$ morphological parameters. Note the dramatic reduction of complexity 60.000 to 180 by moving from full digital to morphological representation.

In ambulatory ECG measurements, the QRS complex clusters are extracted and visualized, furthermore, the trend of the heart rate (HR), the frequency of the different ectopic beats, and the trend of ST segment depressions are plotted. Similar features are represented in intensive care unit reports as well as during surgery. The recurrent problem in ECG interpretation is the qualitative analysis of the QRS complex. The analysis might be very simple, such as the recording of the onset of the QRS complex, or the detection if the support of the QRS complex has widened. A deeper study may reveal phenomena such as the morphological variation of the QRS complexes during arrhythmias.

Our particular interest is in high-resolution ECG, e.g., late potentials or notches and slurs of QRS characterizing fine details of intraventricular depolarization. In high-resolution ECG, the data that we use are the values of the ECG samples. However, in our applications, the meaningful diagnostic information is hidden in the low amplitude, subnoise level, high-frequency details of the ECG signal. In these applications, data should be sampled at no less than 1000 Hz and they are converted to digital signals with at least 12-bit precision. In late potential measurements, most frequently the bipolar X , Y , Z leads are used. To characterize the bioelectrical tissue properties in a more extensive way, the above philosophy has to be extended to the field

of body surface potential mapping as well. In all of these applications, adequate noise reduction is crucial for analysis.

In these applications, data should be sampled at no less than 1000 Hz and they are converted to digital signals with at least 12-bit precision. Most frequently, the bipolar X , Y , Z leads are used in the studies. Adequate noise reduction is crucial for analysis. In practice, mainly the synchronized signal averaging technique is used. According to this approach a new beat should be first aligned to previous beats (template). The program should exclude ectopic or grossly noisy beats. If cross-correlation is used for alignment, the correlation should be performed on the fastest changing part of the QRS at least on a segment of 40 ms. In this paper, illustrative realistic computations are based on Simpson-type X , Y , and Z records.

II. DISTORTION MEASURES

A key problem in generating a set of characteristic ECG-signals, taken of a single patient over a longer period of time, is to define an appropriate metrics in the signal-space and translate this into a definition of metrics in the data-space. Let s be a digital signal obtained by sampling the analog ECG signal of a single heart-cycle. In what follows, an ECG signal is always identified by s unless otherwise stated. The simplest definition of the distance or the *measure of distortion* between s and one of its possible representatives v is

$$\rho(s, v) = (s - v)^T \Sigma^{-1} (s - v).$$

Here, the diagonal matrix $\Sigma = \text{diag}(d_1, \dots, d_m)$, with $d_t > 0$, $t = 1, \dots, m$ takes into account the variability of the ECG signal at different phases of the heart cycle. Thus, in defining similarity of ECG signals, we would, e.g., require higher precision for the P -waves and the T -waves than for the QRS segment. Accordingly, we would use larger weights for samples taken of the P -waves and the T -waves than for the QRS segment. The weighting matrix could also be used to isolate some special features of interest. Initially, we suggest to choose the weights d_t according to the advice of practitioners. For $\Sigma = I$, the identity matrix we get the standard Euclidean distance. Assuming that the Euclidean length of the signals do not differ too much from each other, the above *individual* distortion-measure is approximately equivalent to

$$\frac{\text{corr}(s, v) = (s - \bar{s})^T (v - \bar{v})}{|s||v|}$$

where (\bar{s}, \bar{v}) denote sample means over the cycle and $|s|, |v|$ denote the Euclidean length.

Assume now that ECG signals are represented by morphological parameters, rather than digital signals obtained by sampling. Let these morphological parameters be denoted by ψ . Let the signal s defined above be obtained by the reconstruction procedure $s = F(\psi)$. Then, the representative signals will also be determined by similar morphological parameters, say θ . The distortion-measure, say σ in the space of morphological parameters is given by

$$\sigma(\psi, \theta) = \rho(F(\psi), F(\theta)).$$

III. VECTOR QUANTIZATION

The problem of generating typical signals has been extensively considered in communication theory, and in speech-processing in particular. A basic method for this is a procedure called vector quantization. In vector quantization, we represent vector-valued random variables with a continuous distribution by a small number of prototypes so that an overall distortion measure is minimized. Analogous procedures are also used in cluster analysis (cf. [6]). In speech-processing, the number of classes, which is equal to the number of phonemes, is fixed in advance. For a recent survey of VQ and related procedures see [13].

Vector quantization in its classical form is defined in terms of an Euclidean-metric as follows (cf. [7] and [8]). We are given a data-set $s_i, i = 1, \dots, N$ in say \mathbb{R}^p that has to be classified into d classes. For this purpose, we choose d representatives $v_i \in \mathbb{R}^p$, which are also called centers or prototypes. Note that the representatives $v_i = q(s_i, v)$ are not necessarily members of the data set. Vector quantization is a mapping q from the data set into the set of representatives. Thus, a data-point s will be mapped to $v = q(s)$, where $v = v_i$ for some $i = 1, \dots, d$. In classical vector quantization, we use the distortion measure $\rho(s, v) = (s - v)^T \Sigma^{-1} (s - v)$. Then, the overall measure of goodness of the quantization procedure is given by

$$D = D(v, q) = \frac{1}{N} \sum_{i=1}^N \rho(s_i, q(s_i)), \quad q(s_i) \in \{v_1, \dots, v_d\}. \quad (1)$$

The problem is then to minimize D with respect to the d centers and the mapping q .

It is easy to see that for a fixed set of representatives (v_i) the cost function is minimized if each data-point s is mapped to one of its *nearest neighbor* v , defined by the relation $q(s, v) \leq q(s, v_i)$ for all $i = 1, \dots, d$. The application of this rule then defines a partition of the data space into d classes. Using the nearest neighbor rule, the overall distortion will depend only on the vector v , and we write $D(v, q) = D(v)$. On the other hand, take any fixed partition of the data-space into classes $C_k, k = 1, \dots, d$. Let the signals belonging to class C_k be represented by a yet unspecified v_k . Then, the class-distortion measure for class C_k is defined by

$$D_k = D_k(v_k) = \frac{1}{|C_k|} \sum_{s_i \in C_k} \rho(s_i, v_k) \quad (2)$$

where $|C_k|$ denotes the cardinality of the set C_k . For fixed k , minimizing D_k with respect to v_k , we get the optimizing values

$$v_k = \frac{1}{|C_k|} \sum_{s_i \in C_k} s_i. \quad (3)$$

These simple observations justify a common iterative procedure called the k -means algorithm or Lloyd's method (cf. [6] and [7]). In its off-line version, the data-set or N is fixed. In each step of the iteration, we define a classification of the data-set using the nearest neighbor rule based on the latest set of centers, which is then updated using (3). The on-line version of this procedure is obtained if we allow the data set to be extended in each step of the iteration.

Vector quantization in its classical form is not directly applicable to ECG-analysis, in fact, there is a need to redefine the

class-distortion measure. The definition of class-distortion measure given in (2) is motivated by a probabilistic model. For monitoring purposes, this is not suitable since it is not sensitive to infrequent untypical values which may contain vital medical information. Therefore, we choose to use a *new class-distortion measure* for a given class C_k as follows:

$$D'_k = D'_k(v_k) = \max_{s_i \in C_k} \rho(s_i, v_k). \quad (4)$$

The center of a given finite data set in C_k will be defined as the solution of the problem

$$\min_{v_k} D'_k(v_k) \quad (5)$$

and will be called the *min-max center*. An additional complication arises if we use morphological parameters to define a distortion measure.

Furthermore, in contrast to vector quantization, the number of classes is not fixed in advance in our application. Rather, we fix a tolerance or accuracy, say ε , by which the center v_k must represent the signals in class C_k . If the class-distortion measure exceeds the prescribed threshold value for some k , then we increase the number of classes. The tolerance is chosen basically by medical experience.

It follows, that in any iterative method the exact determination of the minimum of $D'_k(v_k)$ is not necessary, it is sufficient to find just one v_k for which $D'_k(v_k) \leq \varepsilon$. The modification of Lloyd's method is now straightforward. At any time, we define a classification using the nearest neighbor rule based on the latest set of centers v_1, \dots, v_k . If for some k we have $D'_k(v_k) > \varepsilon$, then we solve the minimization problem (5) to find a solution of the feasibility problem $D'_k(v) \leq \varepsilon$. If this problem has no solution then we increase the number of classes: all the outliers, i.e., all the signals that are represented by an accuracy worse than ε , are collected to form one or more new classes.

When $\rho(s, v) = (s - v)^T \Sigma^{-1} (s - v)$, we have

$$D'_k = D'_k(v) = \max_{s_i \in C_k} (s_i - v)^T \Sigma^{-1} (s_i - v).$$

The distance-function is convex in v , hence, its maximum over i , for $s_i \in C_k$ is also convex in v . However, we face a new problem: the function $D'_k(v)$ is not differentiable. Although min-max problems are standard problems for *nonsmooth* optimization (cf. [9, Problem (2.2)]), it is known that when the cardinality of C_k is large N , nonsmooth techniques may break down. On the other hand, as pointed out above, optimality can be relaxed to feasibility. In either way, the bottleneck of the proposed algorithm is the computation of class centers. This problem will be considered in more generality in Sections IV–VI.

IV. SMOOTHING BY RANDOMIZATION

We consider the problem of computing the min-max class centers. We propose to solve this in a suboptimal sense by using a smooth approximation. Let $L(\theta), \theta \in \mathbb{R}^p$ be a general, nondifferentiable, but Lipschitz-continuous cost function, with a Lipschitz-constant K . Assume that $L(\theta)$ has a local, *strictly nonsmooth minimum* at θ^* in the sense that for some $\delta, c > 0$, $|\theta - \theta^*| \leq \delta$ implies

$$L(\theta) - L(\theta^*) \geq c|\theta - \theta^*|. \quad (6)$$

For the minimization of $L(\theta)$, a number of methods are available in the theory of nonsmooth optimization (cf. [9]). However, their efficiency depends considerably on the complexity of the function $L(\theta)$ itself. Here, we develop a new method which seems to be applicable for problems of high complexity.

We generate a sufficiently smooth approximation $\bar{L}(\theta)$ to $L(\theta)$ by locally averaging it, which is implemented by a suitable randomization technique. Let $h(\theta)$ be a sufficiently smooth density function with bounded support, say $h(\theta) = 0$ for $|\theta_i| \geq \delta_m$ for all $i = 1, \dots, p$, with some $\delta_m > 0$. Then, the approximation of $L(\theta)$ will be defined as

$$\bar{L}(\theta) = \int L(\theta + \delta\theta)h(\delta\theta)d\delta\theta. \quad (7)$$

The value of the approximating function can be estimated for each θ as follows: let $\delta\theta$ be a random vector having density function $h(\theta)$. Then, we can write

$$\bar{L}(\theta) = EL(\theta + \delta\theta).$$

If δ_m is small then the probability distribution defined by $h(\theta)$ is concentrated around zero and, hence, $\bar{L}(\theta)$ will be a good approximation of $L(\theta)$. Indeed, it is easy to see that $|L(\theta) - \bar{L}(\theta)| \leq K\delta_m$. On the other hand, since $L(\theta)$ has a strictly nonsmooth minimum at θ^* , it can be shown that for any fixed small $\delta_0 > 0$ the approximating function $\bar{L}(\theta)$ has a local minimum inside the sphere of radius δ_0 around θ^* , if $\delta_m < c\delta_0$ with some fixed $c > 0$. Note that if $L(\theta)$ is convex, then $\bar{L}(\theta)$, as a mixture of convex functions is also convex.

The function $\bar{L}(\theta)$ is not computable explicitly, however, we can easily generate a Monte Carlo estimate $\hat{L}_n(\theta)$ of $\bar{L}(\theta)$ for any fixed θ at any time n . For this purpose, take a random perturbation $\delta\theta_n$ with density function $h(\theta)$ and set

$$\hat{L}_n(\theta) = L(\theta + \delta\theta_n). \quad (8)$$

The difficulty in minimizing $\bar{L}(\theta)$ is that there is no direct way to estimate the gradient of $\bar{L}(\theta)$. Thus, we have to resort to numerical differentiation using the noise-corrupted function values $\hat{L}_n(\theta)$. This can be done in a very effective manner, using a simultaneous random perturbation of the parameters (cf. [1]), which will be described in Section V.

V. THE SPSA METHOD

Let k denote the iteration time. Then, at time k we take a random perturbation vector

$$\Delta_k = (\Delta_{k1}, \dots, \Delta_{kp})^T$$

where Δ_{ki} is a double sequence of independent identically distributed, symmetrically distributed, bounded random variables satisfying certain technical conditions (cf. [1, Section III]). A standard perturbation that will be used is the double sequence Δ_{ki} with $P(\Delta_{ki} = +1) = 1/2$, $P(\Delta_{ki} = -1) = 1/2$ where P denotes the probability-measure.

Now let $0 < c_k \leq 1$ be a fixed sequence of positive numbers. Let $D_0 \subset \text{int}D$ be a compact, convex domain, the technical conditions for which are specified in [14]. At time k and at position $\theta \in D_0$, we take two measurements

$$\hat{L}_k^+(\theta) = \hat{L}_{2k-1}(\theta + c_k\Delta_k) \text{ and } \hat{L}_k^-(\theta) = \hat{L}_{2k}(\theta - c_k\Delta_k).$$

Then, the i th component of the gradient estimator at time k and $\theta \in D_0$ can be written as

$$H_i(k, \theta) = \frac{1}{2c_k\Delta_{ki}}(\hat{L}_k^+(\theta) - \hat{L}_k^-(\theta)). \quad (9)$$

We define $H = (H_1, \dots, H_p)$. In choosing c_k , there is a tradeoff between bias and variance. Large values of c_k induce significant numerical errors in approximating the directional derivatives by finite-difference quotients. On the other hand, small values of c_k would magnify the effect of the noise that is present in $\hat{L}_k^+(\theta)$ and $\hat{L}_k^-(\theta)$. A careful analysis shows that the best choice for c_k is $c_k = c/k^{1/6}$ (see, e.g., [14]). Then, the truncated *SPSA method* is defined as a stochastic gradient method using the gradient estimates above and the resetting mechanism given below.

Let a_k be a fixed sequence of positive numbers with a_k denoting the stepsize at time k . A standard choice for is $a_k = a/k$ with $a > 0$. This choice is justified by the theory of stochastic approximation: the simplest stochastic approximation scheme is the recursive computation of the mean of a sample, in which the step-size $1/k$ shows up in a trivial manner. Let an initial estimate $\hat{\theta}_0$ be given. We generate a sequence of estimated parameters, denoted by $\hat{\theta}_{k+1}$, $k = 0, 1, \dots$. At time k we first define a tentative value

$$\hat{\theta}_{k+1-} = \hat{\theta}_k - a_{k+1}H(k+1, \hat{\theta}_k).$$

If $\hat{\theta}_{k+1-} \in D_0$, then we set $\hat{\theta}_{k+1} = \hat{\theta}_{k+1-}$. Otherwise, we perform a *resetting*: we define $\hat{\theta}_{k+1} = \hat{\theta}_0$.

The advantage of the SPSA method over other stochastic minimization procedures, such as Kiefer–Wolfowitz’s method, is that only two measurements/iteration will be required. Computational evidence shows that the method yields a good suboptimal solution fast up to dimensions in the range of several hundreds variable. Various aspects of convergence have been established in [1], [14], and [15]. In addition, higher order SPSA methods with improved rate of convergence have been developed in [14].

The performance of the SPSA method as applied to nonsmooth problems is also influenced by the choice of the smoothing parameter δ_m . The asymptotic results of [1] are applicable for fixed smoothing parameter, but for decreasing δ_m we have a transient behavior that has not yet been analyzed.

VI. COMPUTATIONAL EXPERIMENTS

Two ECG records, taken from a healthy subjects and a patient with frequent ectopic beats were analyzed. These two persons are referred to as Patient 1 and Patient 2. In the measurements, the orthogonal X, Y, Z leads suggested by Simpson in [2] for ventricular late potential studies were used. The length of the records were 2.5 min, and from all the technical aspects the measurements were conform with the recommendations formulated in [16].

In the signal preprocessing, at first all the QRS complexes were identified, with the method of Pan and Tompkins [17]. Subsequently, in each QRS complex the time instant of the steepest negative slope in one of the filtered Simpson channel was assigned as temporal fiducial point of the QRS complex. Because the quality of the averaged heart cycle is heavily influenced by any jitter in the fiducial point localization, a fine

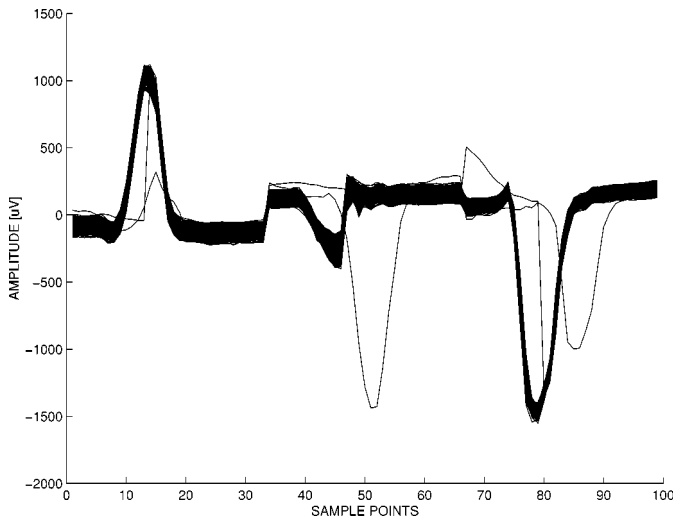


Fig. 1. The ECG signals of Patient 1. Most of the complexes have a similar shape.

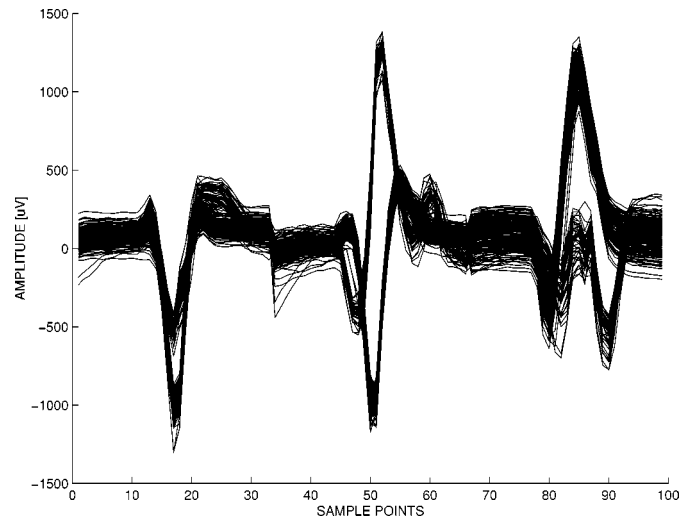


Fig. 3. The ECG signals of Patient 2. Clearly, two types of QRS waveforms are present in the record.

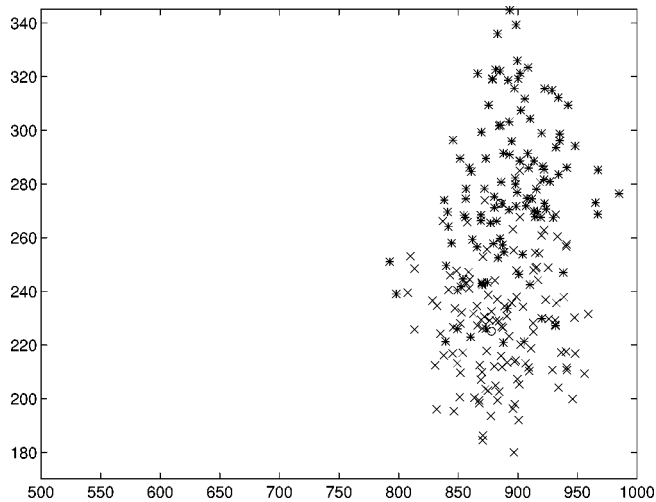


Fig. 2. Projected data points and centers of the two covering spheres for Patient 1. Data points corresponding to the two spheres are indicated by different symbols.

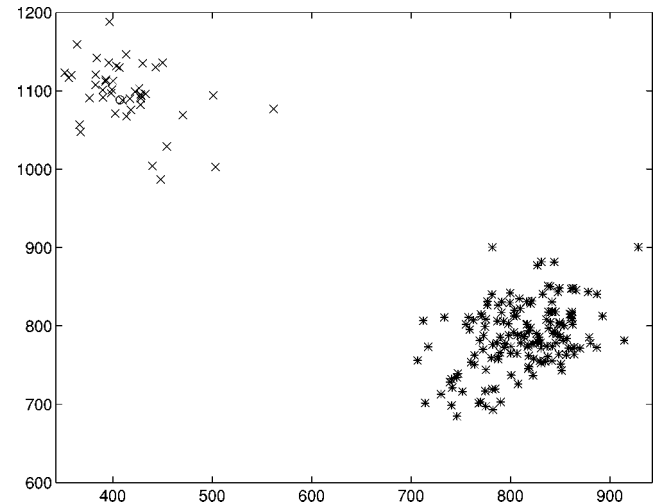


Fig. 4. Projected data points of Patient 2. Clearly, data belong to two different subgroups.

adjustment of the fiducial point position is usually required. The routine procedures use correlation coefficients to this purpose.

In Experiment 1, the three leads for the same heart cycle were represented by 33 numbers/lead, the middle point representing the fiducial point of the cycle. Thus, our data sets consisted of points in 99 dimension. We can reduce the dimension by choosing the signal-values at the fiducial point and four points before and after it.

Thus, we get a $3 \times 9 = 27$ -dimensional measurement vector. The number of cycles or data points were 242 and 204, respectively. In Figs. 1 and 3, we plot superimposed Simpson-type leads of the two subjects measured. In each graph, the tracing of the first third represents the X leads, the second third the Y leads, and the last third the Z leads. The superimposed signals were time-aligned by the fiducial points of the QRS complexes.

These data sets are analyzed using our classification method. Our results revealed that the new method correctly recognized

ectopic beats and artifacts, and the majority (normal) heart cycles.

In the case of Patient 1, we first identified two outliers, then we fitted two spheres to the remaining 240 points data-points (cf. Fig. 2). The results for Patient 2 are given on Fig. 4. In this case, we have no outliers.

As we can see, in all of our ECG examples the QRS patterns were clustered in two groups. For Patient 2, the clusters separated significantly. A deeper study on the separation and its medical utility seems to be justified but it is beyond the scope of this paper.

In Experiment 2, we compare the min-max classification method for QRS patterns with the widely used correlation procedure. Let us repeat that one objective of the classification to generate reliable typical signals.

A critical issue in the synchronization and noise reduction of quasiperiodic biological signals is the selection of signals that will be used to form an averaged signal. The samples meet all international criteria [16].

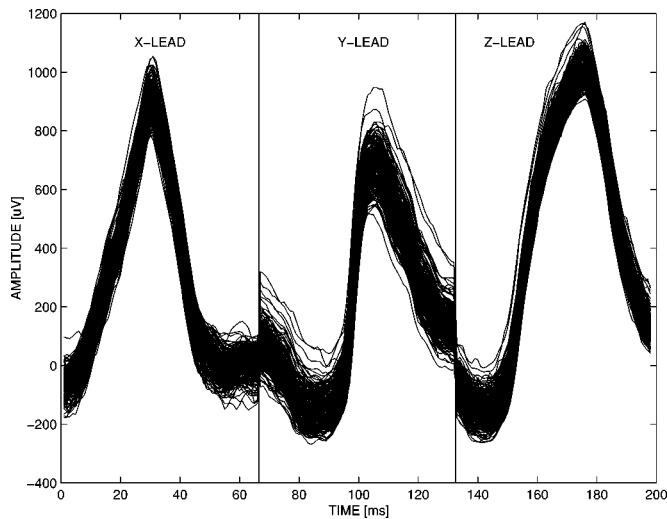


Fig. 5. Superimposed cycles of a typical ECG record for comparison.

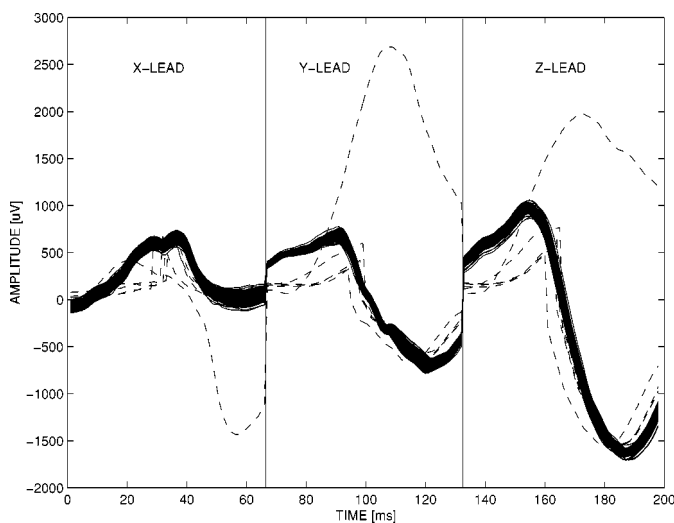


Fig. 6. Superimposed cycles of an ECG record with a few extra beats, used for comparison in Section VI.

For comparison, we use two records (see Figs. 5 and 6). The first record is an “excellent” sample, (all cycles are very similar), while the second record contains five ventricular extra-beats. There is a time-window around the fiducial-point, lead by lead, with a length of 66 ms.

For the min-max criterion, the signal obtained during the time-windows of the three leads are concatenated. Thus, a single cycle will be represented by $3 \times 66 = 198$ sampling point, i.e., one cycle is a point of the 198-dimensional space. These strings are shown in Figs. 5 and 6.

In the correlation method, first a reference cycle is calculated for all the time- windows lead by lead. Subsequently, the cycles are compared: if a cycle on one of the leads has a correlation coefficient smaller than 0.98, we drop that cycle from all three leads. The threshold value 0.98 meets to international standards [16].

In Fig. 7 the ECG-record of Fig. 5 is analyzed. The correlation coefficient of each cycle is above 0.98, hence, no cycle will be qualified as an outlier. On the other hand, the homogeneity of the record can be improved using min-max classification with two

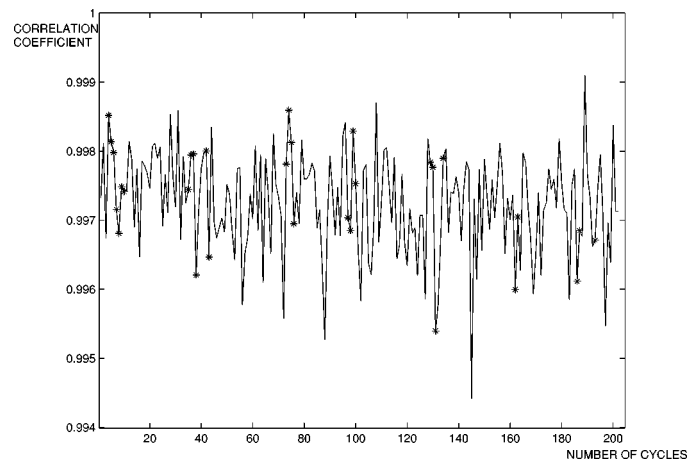


Fig. 7. Correlation coefficients of the first ECG record; the signals marked with * were deleted by the SPSA method.

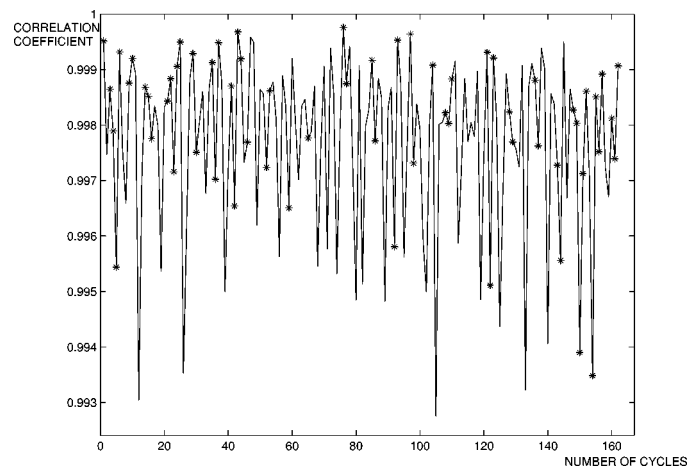


Fig. 8. Correlation coefficients of the second ECG record; the signals marked with * were deleted by the SPSA method.

classes, and removing the data belonging to the class containing less samples. The deleted data-points are marked by *.

In Fig. 8, the ECG record on Fig. 6, containing five extra beats, represented by dotted lines, is analyzed. The correlation method detects these extra beats, but no other cycles are deleted.

To get a more homogenous record, we apply min-max classification with two classes on these preclassified data, as in the previous experiment. An interesting feature of this procedure is that the cycles that closely precede or follow the extra beats have been automatically deleted, although these extra beats are not present in the preclassified data-set. This is in agreement with the common practice that the mentioned cycles are deleted manually when the correlation method is applied.

The conclusion of Figs. 7 and 8 is that the set of QRS complexes that are similar in terms of the correlation method coefficients may eventually be separated by the SPSA method. A careful comparison using synchronized averages and second derivatives results in two representative signals that are slightly shifted and also different in fine details of their pattern. The biophysical background of the second derivatives suggest that pattern changes might be due to slight differences in local intraventricular activation propagation velocities, causing shifts in the timing of collisions of activation wavelets originated in distinct

points of the Purkinje fiber-endocardial interface. We have to emphasize that the above interpretation needs further confirmation, though previous numerical modeling studies support our hypothesis [18].

These findings indicate that min-max classification may give useful complementary insight to detect minor morphological differences. An advantage of the min-max method is that it allows the user to assign different weights to different coordinates, or even morphological parameters. Different coordinates may carry different biomedical content based on the expertise of cardiologists. Tuning these weights is an additional degree of freedom in the hands of the cardiologist to detect specific abnormalities.

VII. DISCUSSION

We have proposed a new procedure for the classification of ECG signals taken over a longer period of the same patient. The procedure is obtained by modifying the well-known vector-quantization procedure. First, the center of a class is defined via a min-max procedure, rather than by averaging signals within the given class. Second, the number of classes is not fixed in advance, instead an upper bound or tolerance for the class distortion measure is given. The procedure can be used for monitoring purposes: an alarm signal will be generated if untypical signals are recorded.

A main characteristic of the new procedure is its enhanced sensitivity to untypical signals and outliers. The price of this practical advantage is increased computational complexity. Class-centers are defined in terms of a nonsmooth optimization problem, which is solved by a modification of the so-called SPSA method. The basic ingredients of the relevant theory and the results of computational experiments are given on both simulated and real data. The latter are ECG signals that have been preprocessed for the detection of low-level ECG components, e.g., late potentials. Parts of this paper have been published in abridged form in the conference-papers [19], [20].

The proposed approach can be extended in a number of ways. A primary design parameter of the method are the weights d_t . We assumed that these are selected on the basis of medical expertise. However, an equally sensible procedure is to allow them to be tuned within some limits. Another interesting problem is to see if the classification is stable when the level of tolerance is adjusted.

ACKNOWLEDGMENT

The authors would like to thank J. C. Spall and J. L. Maryak for initiating and cooperating in this research, and G. Bortolan for helpful discussions.

REFERENCES

- [1] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 332–341, Mar. 1992.
- [2] M. Simpson, "Use of signals in the terminal QRS complex to identify patients with ventricular tachycardia after myocardial infarction," *Circulation*, vol. 64, pp. 235–242, 1981.
- [3] E. Berbari, "High resolution electrocardiography," *CRC Crit. Rev. Biomed. Eng.*, vol. 16, pp. 67–103, 1988.

- [4] Z. W. L. R. EJ, C. JP, K. JA, and M. AJ, "ECG features of microvolt T-wave alternans in coronary artery disease and long qt syndrome patients," *J. Electrocardiol.*, vol. 31, no. (Suppl.), pp. 114–120, 1998.
- [5] F. Greensite, "Well-posed formulation of the inverse problem of electrocardiology," *Ann. Biomed. Eng.*, vol. 22, pp. 172–183, 1994.
- [6] M. R. Anderberg, *Cluster Analysis for Applications*. New York: Academic, 1973.
- [7] Y. X. D. Huang and M. Jack, *Hidden Markov Models for Speech Recognition*. Edinburgh, U.K.: Edinburgh Univ. Press, 1990.
- [8] R. Gray, "Vector-quantization," *IEEE Acoust. Speech, Signal Processing Mag.*, pp. 4–29, 1984.
- [9] C. Lemarechal, "Nondifferentiable optimization," in *Handbooks in Operations Research and Management Science*. Amsterdam, The Netherlands: Elsevier Science, 1989, pp. 529–572.
- [10] S. Jalaeddine, C. Hutchens, W. Coberly, and R. Strattan, "Compression of holter ECG data," *Biomed. Sci. Instrum.*, vol. 24, pp. 35–45, 1988.
- [11] P. W. Macfarlane and T. D. Veitch Lawrie, *Comprehensive Electrocardiology*. Oxford, U.K.: Pergamon, 1989.
- [12] J. L. Willems, P. Arnaud, J. H. Van Bommel, R. Degani, P. W. Macfarlane, and Chr. Zywiets, "Common standards for quantitative electrocardiography: Goals and main results," *Meth. Inform. Med.*, vol. 29, pp. 263–271, 1990.
- [13] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: Applications and algorithms," *SIAM Rev.*, vol. 41, pp. 637–676, 1999.
- [14] L. Gerencsér, "Rate of convergence of moments for a simultaneous perturbation stochastic approximation method for function minimization," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 894–905, May 1999.
- [15] H. Chen, T. Duncan, and B. Pasik-Duncan, "A stochastic approximation algorithm with random differences," in *Proc. 13th Triennial IFAC World Congr.*, J. Gertler, J. Cruz, and M. Peshkin, Eds., San Francisco, CA, 1996, pp. 493–496.
- [16] G. Breithardt, M. Cain, N. El-Sherif, N. Flowers, V. Hombach, M. Janse, M. Simpson, and G. Steinbeck, "Standards for analysis of ventricular late potentials using high resolution or signal-averaged electrocardiography," *Eur. Heart J.*, vol. 12, pp. 473–480, 1991.
- [17] J. Pan and W. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp. 230–236, 1985.
- [18] G. Kozmann, Z. Cserjés, M. Tysler, M. Tinova, M. Turzova, and V. Szathmáry, "Feasibility of a predictive algorithm for identifying characteristic events of cardiac activation: Results of a model study," in *Computers in Cardiology*. Piscataway, NJ: IEEE Press, 1994, pp. 801–804.
- [19] L. Gerencsér, G. Kozmann, and Z. Vágó, "Non-smooth optimization via SPSA," in *Proc. Conf. Mathematical Theory of Networks and Systems, MTNS'98*, Padova, Italy, pp. 803–806.
- [20] —, "SPSA for nonsmooth optimization with application in ECG analysis," in *Proc. Conf. Decision and Control*, vol. CDC 37, 1998, pp. 3907–3908.



László Gerencsér received the M.Sc. and doctorate degree in mathematics from the Eötvös Loránd University (ELTE), Budapest, Hungary, in 1969 and 1970, respectively. In recognition of his research activity, he received the title Doctor of the Hungarian Academy of Sciences in 1999.

Since 1970, he has been with the Computer and Automation Institute of the Hungarian Academy of Sciences, MTA SZTAKI, Budapest, where he currently heads the Applied Mathematics Laboratory. He held a one-year visiting position at the

Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden, in 1986. From 1988 to 1991 he was a Visiting Professor with the Department of Electrical Engineering, McGill University Montreal, QC, Canada. From 1997–2001 he held a Széchenyi Professorship with the Eötvös Loránd University, Budapest. His main recent research interests include: hidden Markov-models, statistical theory of stochastic systems, stochastic adaptive control, stochastic complexity, randomization methods in optimization, and financial mathematics. He has been a member of the Conference Editorial Board of the IEEE Control Systems Society since 1998. Since its foundation, he has been an Associate Editor for the *Journal of Mathematical Systems, Estimation and Control*. Since 1998, he has been an Associate Editor for the *SIAM Journal of Control and Optimization*, and since 2001, he has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.



György Kozmann received the M.Eng. degree in electrical engineering from the Technical University of Budapest, Budapest, Hungary, in 1964, and the candidate of technical sciences and the D.Sc. degrees from the Hungarian Academy of Sciences in 1981 and 2001, respectively.

He is Professor and Head of the Department of Information Systems, University of Veszprém, Veszprém, Hungary, and Scientific Adviser and Head of Department for Bioengineering at the Research Institut for Technical Physics and Materials Science,

Budapest.

Dr. Kozmann is President of the Biomedical Division of the J. von Neumann Society, and Editor-In-Chief of the Hungarian periodical *Information and Management in Healthcare*.



Kristóf Haraszti received the M.Eng. degree in electrical engineering from the Budapest University of Technology and Economics, Hungary, in 2001. He is working towards the Ph.D. degree at the University of Veszprém, Veszprém, Hungary. His Ph.D. degree project is on the early detection of increased malignant arrhythmia vulnerability through bioelectrical measurements.

He is working at the Department for Bioengineering at the Research Institut for Technical Physics and Materials Science, Budapest.



Zsuzsanna Vágó received the M.Sc. and Ph.D. degrees in mathematics from the Eötvös Loránd University, Budapest, Hungary, in 1982 and 1996, respectively.

She is a Senior Researcher at MTA SZTAKI (the Computer and Automation Research Institute of the Hungarian Academy of Sciences), Budapest, and an Associate Professor with the Pázmány Péter Catholic University, Budapest. Her main research interest include statistical theory of stochastic systems, randomization methods in optimization and

financial time-series.