DOI 10.5162/sensor2015/P6.1

Reference-less Human Motion Recognition using MEMS-based inertial motion sensors and stochastic signal modelling

Mustafa Kamil[1], Markus Haid[1], Thitipun Chobtrong[1], Ersan Günes[1], Patrick Abrante-Perez[1], Nick Berezowski[1]
[1] CCASS – Competence Center for Applied Sensor Systems,
64295 Darmstadt, Germany, www.ccass.h-da.de
mustafa.kamil@h-da.de

Abstract:

Human Motion is a highly variable and multidimensional form of displacement and rotation series in space performed by multiple parts of a moving body, i.e. different muscles, bones and joints working together. As adult humans have mastered to optimize different movements in early childhood (learning from other people and/or from own mistakes), these movements seem obvious to them in everyday life and hence evoke no need for further query or perfection. In professional sports or in applications of rehabilitation and advanced training a reliable possibility of computer-assisted motion analysis and validation can be a key for optimized training procedures and success measurement. The present work shows the latest research results performed at the CCASS aiming for providing a framework for reference-less human motion analysis and validation using low-cost inertial motion sensors and a light-weight, full-body mutli-sensor suit. The developed algorithms base on the theory of Hidden Markov Models and on stochastical modelling of human motion using Markov chains. In the present paper the motion recognition concept will be explained as well as the model definition, the feature selection and the validation results will be discussed. Ultimately, impressions from the sensor suit development and the future work will be given.

Key words: Inertial Measurement Unit (IMU), Micro-electro-mechanical systems (MEMS), Motion Recognition, Motion Capture, Medical Sensing, Sport Sciences, Multisensor-System

Introduction

Aiming for a novel and low-cost computer-assisted motion monitoring and optimization concept, the present work starts from the idea of dealing with human motion as a series of displacements and rotations performed by a moving body. In a more detailed consideration, motion is a holistic concept consisting of different individual components which take part in performing a specific desired action. Being able of identifying these components as well as any failure or disturbances in an observed motion sequence could be employed for highly efficient motion guidance and optimized training for several applications given in professional sports, medicine, health care and any other conceivable use case.

The current state of the art in full-body motion recognition is mainly dominated by machine vision and marker-based systems. Especially marker-based approaches can offer a highly accurate motion capturing but they also suffer from several drawbacks, s.a. shading effects, additional infrastructure and a high installation effort. An elegant approach to generate motion-sensitive measurements without the need of external signals or devices can be provided by inertial navigation systems.

Inertial Navigation Systems

The principle of inertial navigation systems (INS) is based on the measurement of motion using the inertia of a sensor-integrated mass in the case of its acceleration. INS are generally capable of measuring object motion without needing neither external signals or aiding infrastructure and can be realized using each three turn rate and acceleration sensors forming an orthogonal assembly in the three-dimensional space. Having the sensitive sensor axes in the directions of a Cartesian axis frame, it is possible to measure any force acting on an observed object in rotational or linear direction and hence to cover all of the six degrees of freedom a moving body can have in three-dimensional space (cf. figure 1).

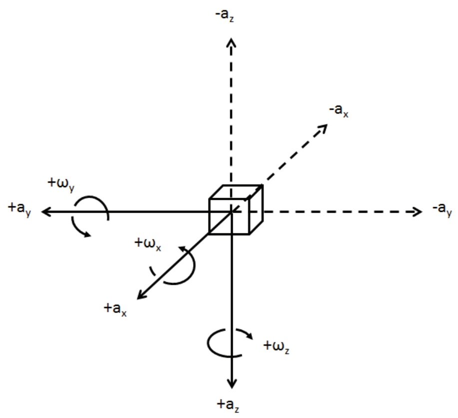


Fig. 1. Schematic representation of inertial motion measurement using an orthogonal assembly of turn rate and acceleration sensors

A decisive challenge given for low-cost INS in motion tracking and recognition applications is, that micromechanical inertial sensors suffer from bias errors of which especially the random part can't be compensated trivially. This results in a loss of accuracy over time due to the numerical integration of biased acceleration and turn rate signals, causing rapidly growing navigation errors. As the velocity, position and turn angle values are only accurate over short periods of time, low-cost INS are regarded as short-term stable systems [1,2]. The main task in the present approach is therefore given in finding a way to recognize and classify motion while preventing numerical integration of low-cost INS signals. The result shall be applied for human motion recognition through using multiple low-cost INS operating simultaneously inside a full-body multisensor suit (cf. figure 2).

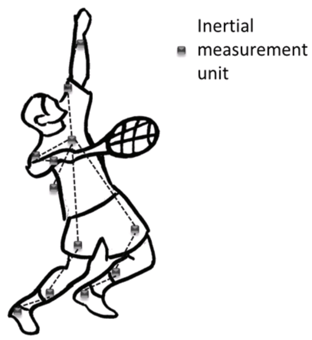


Fig. 2. Concept of a full-body multisensor system containing twelve inertial measurement units for reference-less three-dimensional motion recognition

Motion recognition

The present motion recognition concept comprises an algorithmic framework consisting of several signal processing and recognition algorithms and a user interface consisting of an inertial sensor suit and a data acquisition software interface. The signal processing comprises a feature acquisition part, a motion model definition and optimization part and a final part for generating the recognition results.

Motion modelling

One of the most known applications of modelling "hardly observable" processes aiming for deriving possible explanations for their nature using the process outputs only is given in the automatic recognition of human speech. The present work tries to benefit from the experiences collected in the speech recognition segment and aims for treating motion-generated sensor measurements using the same techniques as for speech signals.

The idea applied in the present approach regards acceleration and turn rate signals generated by a motion process as outputs produced by a natural process, which itself can't be observed directly using low-cost INS. This is the same for speech recognition, as spoken sounds can be sensed well, where the text responsible for their generation stays "hidden". Next to speech recognition, explaining hardly observable processes using the outputs has gained high importance in many modern applications, e.g. in the weather forecast, stock market or genetic research [3].

Powerful mathematical tools for modelling hardly observable processes can be obtained from the theory of Hidden Markov Models (HMMs). HMMs can be understood as a mathematical representation of a finite state machine, of which the state transitions over time depend on probabilities instead of absolute conditions. In an HMM, the states usually correspond to a specific instant of time a process of interest produces an output, that can be sensed using an appropriate measurement.

Both the transitions between the states over time as well as the production of outputs with each new transition correspond to well-defined probabilities, making an HMM generally describing two simultaneously occurring stochastic processes [4-6]. Moreover, an HMM is always subjected to the Markov property, meaning that a future system state is regarded as conditionally independent of the time as well as of any past state, provided that the current system state is given (1st order HMM is considered, cf. equation 1, [4-6]).

The definition of the probability connecting two system states over an instant of time can be defined by equation 1. Next to this definition it must be respected, that all forbidden transitions are assigned to a zero probability and that all outgoing transition probabilities of a single state must sum to one.

$$P\left(q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots\right) = P\left(q_t = S_j | q_{t-1} = S_i\right) = a_{ij} \quad 1 \le i, j \le N \tag{1}$$

The second stochastic process described by an HMM covers the relation between an observed output and its probabilistic subjection to the generating state. Treating this mathematically can be significantly simplified, if the variety of observable outputs can be limited to a finite, discrete number, which in the original process maybe of continuous nature, i.e. their number can be unlimited. The dependency between the system states and the observable outputs, the emission probabilities, can then be described by equation 2. It can be said, that a system going through a certain sequence of states over time must produce a certain sequence of observable outputs with respect to a precisely calculable probability [4-6].

$$P\left(\begin{matrix}\underline{v}_{k,t} | q_t = S_j, q_{t-1} = \\ S_i, \dots, \underline{o}_1, \underline{o}_2, \dots \underline{o}_{t-1}\end{matrix}\right) = P\left(\underline{v}_{k,t} | q_t = S_j\right) = b_j(k) \quad 1 \le j \le N,\ \ 1 \le k \le M \tag{2}$$

Limiting the number of the system outputs requires an appropriate discretization algorithm. The final number of possible outputs (also: codebook size or alphabet) will affect the algorithmic accuracy as well as the computational cost. Therefore, a compromise must be found in order to guarantee sufficient output differentiability while keeping the alphabet size preferably small.

The discretization technique applied in this approach is based on the Euclidean distance method (cf. equation 3, [7]). Moreover, the decision was made for an output alphabet of 26 equally distributed discrete unit vectors pointing to a number of representative directions in the three-dimensional space (cf. figure 3). As a supplementary output, the vector magnitudes were included in the form of a discrete magnitude level scale (4^{th} feature, cf. results).

$$\epsilon_{dist}(\underline{o}_k, \underline{v}_k) = \sqrt{\sum_{k=1}^{M} \left(\underline{o}_k - \underline{v}_k\right)^2} \tag{3}$$

The motion recognition concept is realized by defining a number of independent HMMs, each observing one of the inertial measurement units inside the multisensor suit. Furthermore, these models are redefined every time the motion type is changed, i.e. another application use case is selected.

For the hand wrist measurement unit in the sample use case of a handball throw the corresponding HMM (cf. figure 4) comprises five motion states referring to the stages of an ideal hand wrist motion sequence (cf. figure 5). As the same technique can be repeated for an unlimited number of measurement units inside the sensor suit, the present article will concentrate on the hand wrist HMM.

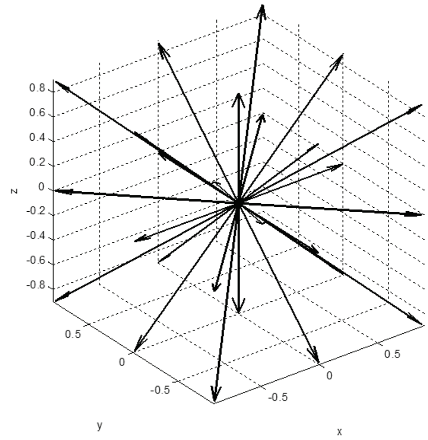


Fig. 3. Reference vectors defining the quantization codebook

After initializing an appropriate HMM (cf. equations 4,5, [4-6,8]) for a selected measurement unit and motion type, it is necessary to provide a sufficient amount of motion patterns performed as perfectly as possible in order to optimize the model parameters for the recognition quality required by the user. While the most classical HMM approaches concentrate on using a large number of data sets in order to reach best possible optimization results, the present work presents the alternative approach of single sample training (cf. section "Single sample training").

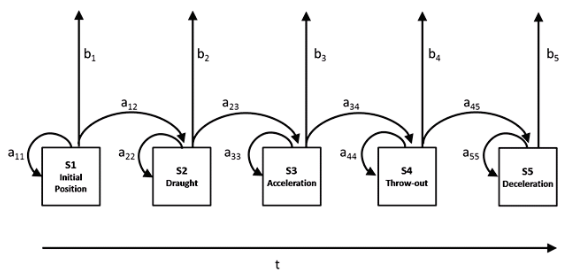


Fig. 4. Straight-forward HMM for the handball use case and the hand wrist sensors [4-6]

DOI 10.5162/sensor2015/P6.1

$$a_{ij} = 0 \textit{ für alle } j < i \textit{ und } j \geq i+1 \quad (4)$$

$$b_j(k) = \frac{1}{M}, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (5)$$

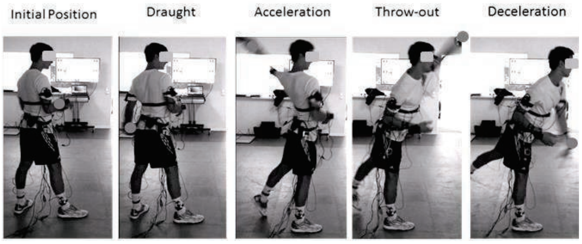


Fig. 5. Five stages representing the throwing hand wrist motion

The model optimization (also: training) can be performed efficiently using well-known and recognized algorithms. One of the most known approaches lies in the expectation maximization technique presented by Baum et al. [9], which requires a number of sub-algorithms for the calculation of different important probabilities inside a given HMM. The first and most important calculation is the probability for a model to generate a given observation sequence and can be calculated using the forward algorithm summarized in the equations 6 and 7. The forward algorithm can be applied to serve the model-based recognition and is essentially required for the parameter optimization [4-6].

$$\alpha_{t+1}(j) = \left[\sum_{i=t}^{N} \alpha_t(i)\, a_{ij}\right] b_j\left(\underline{o}_{t+1}\right), \quad 1 \leq t \leq T-1, \; 1 \leq j \leq N \quad (6)$$

$$P(\boldsymbol{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \quad (7)$$

A complementary approach is the backward algorithm, which specifies a variable calculated backwards in time. The backward variable provides the probability of a state sequence to generate the observation sequence after a specified time step and can be calculated equivalently to the forward variable as summarized by equation 8 [4-6].

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, b_j\left(\underline{o}_{t+1}\right) \beta_{t+1}(j), \quad t = T-1,\; T-2, \dots 1, \; 1 \leq i \leq N \quad (8)$$

Finding the most optimal parameters for a motion HMM can be reached by combining the forward and backward algorithms using a number of ideal motion measurements and the initial model. The idea of Baum et al. lies in calculating the model likelihood to generate a given output sequence, then changing the parameters and checking them for likelihood improvement inside an iterative loop. It goes back to Baum et al. that the observation probability can be locally maximized within a finite number of loop iterations, provided that the training patterns and the initial parameters were chosen properly. This method is also referred to as Baum-Welch algorithm and is summarized in the equations 9-12 [4-6,9].

$$\gamma_t(i) = P(q_t = S_i|\boldsymbol{O},\lambda) = \frac{P(\boldsymbol{O}, q_t = S_i|\lambda)}{P(\boldsymbol{O}|\lambda)} = \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\,\beta_t(i)}, \quad 1 \leq t \leq T, \; 1 < i < N \quad (9)$$

$$\xi_t(i,j) = P\left(q_t = S_i, q_{t+1} = S_j|\boldsymbol{O},\lambda\right) = \frac{\alpha_t(i) a_{ij}\, b_j\left(\underline{o}_{t+1}\right) \beta_{t+1}(j)}{P(\boldsymbol{O}|\lambda)}, \quad 1 \leq t < T, \; 1 < i,j < N \quad (10)$$

$$\tilde{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)}, \quad 1 < i,j < N \quad (11)$$

$$\tilde{b}_{j(k)} = \frac{\sum_{t\,with\ \underline{o}_t=\underline{v}_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}, \quad 1 < j < N, \; 1 \leq k \leq M \quad (12)$$

Model-based recognition

The observation probability calculated in the forward algorithm is a quite qualified measure for comparing a given motion model with an observed output sequence, especially if only one model comes into question. Another, more powerful method is the Viterbi algorithm, which extends the forward algorithm towards finding the most likely state sequence inside the model (cf. equations 13-17, [4-6,10]).

$$\delta_t(i) = \max_{q_1,q_2,\dots q_{t-1}} P\left(\begin{matrix} q_1,q_2,\dots q_t = i, \\ \underline{o}_1,\underline{o}_2 \dots \underline{o}_t \end{matrix}\middle|\lambda\right) \quad (13)$$

$$\delta_t(j) = \max_{1 \le i \le N} \left[\delta_{t-1}(i)\, a_{ij}\right] b_j(\underline{o}_t), \quad 1 \le j \le N,\ 2 \le t \le T \tag{14}$$

$$\psi_t(j) = \underset{1 \le i \le N}{argmax}\left[\delta_{t-1}(i)\, a_{ij}\right], \quad 1 < t \le T,\ \ 1 < j < N \tag{15}$$

$$q_T{}^* = \underset{1 \le i \le N}{argmax}\left[\delta_T(i)\right] \tag{16}$$

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots 1 \tag{17}$$

Single sample training

As discussed above, most HMM approaches rely on the presence of a sufficient amount of training data sets, i.e. on multiply repeated ideal motion measurements for optimizing the model parameters. In practice, however, this is often not the case as the time or the effort required to generate these patterns can be high or the availability of a person capable of performing ideal motion prototypes can be limited, e.g. professional athletes, a trainer, etc.

The single sample approach presented in the present work takes advantage of the discrete reference vectors presented before. The quantization algorithm provides a number of vectors, of which each can be imagined as surrounded be a limited number of neighboring vectors (cf. figure 6). In the present approach, every reference vector is related to eight other vectors by means of a minimum Euclidean distance, so that each vector can be varied through replacement by one of its neighbors. Performing this procedure in each a single segment given in motion sequence (while the other segments stay original) allows for generating a large number of training patterns derived from only one true ideal motion sample. While realizing a certain amount of desirable model tolerance, the effort required for the true data generation stays small and the model stays optimally trainable.

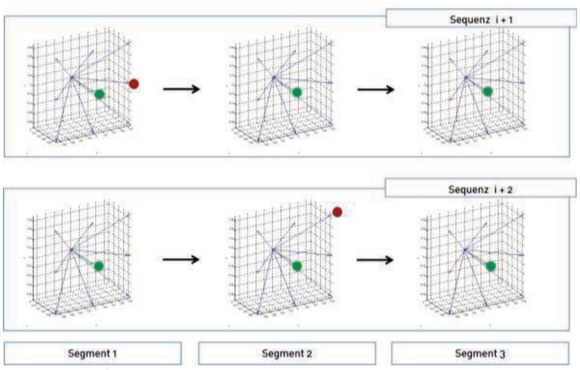


Fig. 6. Single sample training concept

Multisensor suit

The main user interface for the motion recognition is designed to be a close-fitting, full-body sensor suit in which a specific number of individually operating inertial measurement units are be integrated. As the measurement units developed in the present work don't exceed the size of 2 x 4 cm, the final sensor network can be integrated into small, invisible pockets inside the suit while being able of wireless communication with both each other as well as an external processing computer. As a result, the final application is independent of additional infrastructure or installations as well as undisturbed by hardware or cables inside the suit (cf. figure 7).

Fig. 7. Multisensor suit for the motion recognition

System Validation

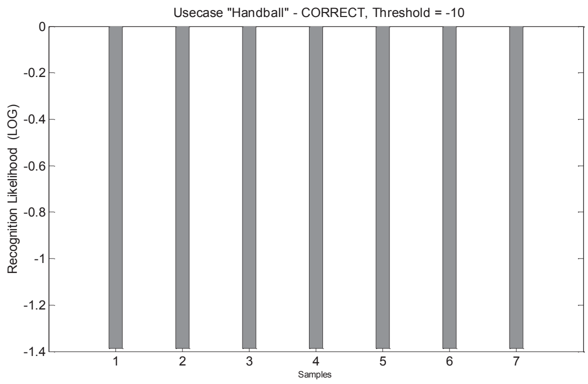
In order to validate the present system stability and reliability under authentic conditions, the decision was made for a real motion capturing work shop with professional athletes. The study participants were equipped with each seven inertial measurement units all over the body, in order to show, that the model-based recognition can be performed for all components of a full-body motion, i.e. the number of measurement units is unlimited.

The participants were asked to repeat certain motion use cases for each a typically wrong performance and for the motion pattern they would regard as ideal. For the hand ball use case referred to above, a five-state HMM (cf. figure 4) was employed to validate the quality of the motion patterns using the unit vector features and alternatively the unit vectors plus the 3D-Vector magnitude. The recognition results for both the three and the four feature experiments are provided in the figures 8 - 11.

Discussion and outlook

As shown by the diagrams in the figures 8 - 11, the motion validation using the present concepts could be performed with sufficiently successful recognition results. It could also be shown, that including a fourth or more features can help to suppress false positive samples, resulting in a higher recognition quality.

Future developments will concentrate on further system tests under stress conditions and on continuous improvement of the discussed algorithms. The usability and the feedback generation shall also be extended and optimized. Moreover, current hardware developments aiming for extremely miniaturized measurement units and highest possible range of wireless data transmission will be continued.



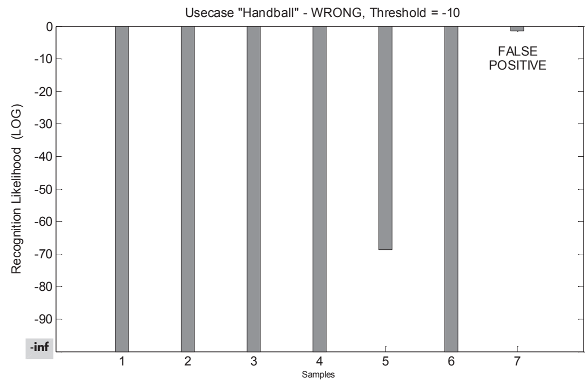
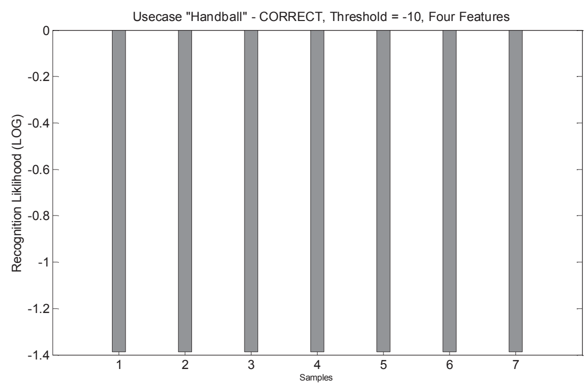


Fig. 8, 9. Recognition Likelihood (LOG) for Usecase "Handball" – CORRECT (fig. 8) and WRONG (fig. 9), ***Threshold = -10***



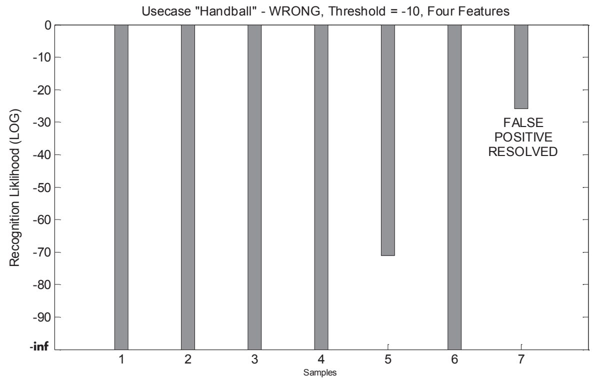


Fig. 10, 11. Recognition Likelihood (LOG) for Usecase "Handball" – CORRECT (fig. 10) and WRONG (fig. 11), ***Threshold = -10****, four features*

References

[1] Barbour, N., Schmidt, G.: Inertial sensor technology trends. IEEE Sensors J. 1(4), 332–339 (2001).

[2] Haid, M.: Verbesserung der referenzlosen inertialen Objektverfolgung zur low-cost Indoor-Navigation durch Anwendung der Kalman-Filterung. Fraunhofer-IRB-Verl., Stuttgart (2005)

[3] Eddy, S.R.: What is a hidden Markov model? Nature Biotechnology 22, 1315 - 1316 (2004).

[4] Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge, Mass (1999)

[5] Pfister, B., Kaufmann, T.: Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer, Berlin, Heidelberg (2008)

[6] Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-sequential images using hidden Markov model. In: 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, USA, 15-18 June 1992, pp. 379–385.

[7] Rizvi, S.A., Nasrabadi, N.M.: An efficient Euclidean distance computation for vector quantization using a truncated look-up table. IEEE Trans. Circuits Syst. Video Technol. 5(4), 370–371 (1995).

[8] Rabiner, L.R., Levinson, S.E., Sondhi, M.M.: On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition. Bell System Technical Journal 62(4), 1075–1105 (1983).

[9] Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Ann. Math. Statist. 41(1), 164–171 (1970).

[10] Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Inform. Theory 13(2), 260–269 (1967).