Abstract—While it is a common knowledge that AC coefficients of Fourier-related transforms, like DCT-II of JPEG image compression, are from Laplace distribution, there was tested more general EPD (exponential power distribution) $\rho \sim \exp(-|x|^\kappa)$ family, leading to maximum likelihood estimated (MLE) $\kappa \approx 0.5$ instead of Laplace distribution $\kappa = 1$ - such replacement gives $\approx 0.1$ bits/value mean savings (per pixel for grayscale, up to $3\times$ for RGB). There is also discussed predicting width (scale parameter) for AC coefficients based on already decoded coefficients (as linear combination of their absolute values), leading to much larger $\approx 0.5$ bits/value mean savings.

Especially for such continuous distributions, there is also discussed quantization approach through optimized continuous quantization density function $q$, which inverse CDF (cumulative distribution function) $Q$ on regular lattice $\{Q^{-1}(\kappa i/N) : i = 1 \ldots N\}$ gives quantization nodes - allowing for flexible inexpensive choice of optimized (non-uniform) quantization - of varying size $N$, with rate-distortion control. Optimizing $q$ leads to significant improvement, however, at cost of increased entropy due to more uniform distribution. Optimizing both turns out leading to nearly uniform quantization here, with automatized tail handling.

Keywords: image compression, quantization, discrete cosine transform, rate-distortion optimization

I. INTRODUCTION

Modern lossy image/video compression is usually based on Fourier-related transforms, especially discrete cosine transform DCT-II used e.g. in JPEG image compression [1]. While DC coefficients describing mean value have completely different behavior, requiring separate treatment usually similar to lossless image compression, the AC coefficients are usually assumed to be from Laplace distribution [2].

This assumption is verified here using more general family: EPD (exponential power distribution) [3], [4]: $\rho(x) \propto \exp(-(|x-\mu|/\sigma)^\kappa/\kappa)$ covering both Laplace distribution for $\kappa = 1$, Gaussian distribution for $\kappa = 2$, and other behaviors of both body and tail of distribution. MLE (maximum likelihood estimation) allows to test if standard $\kappa = 1$ assumption is the proper one, but for AC coefficients it clearly leads to essentially smaller $\kappa \approx 1/2$, as shown if Fig. [1].

Such replacement allows to improve compression ratio by $\approx 0.1$ bits/value, which seems significant as for RGB we get $\approx 0.3$ bits/pixel this way (or less for chroma subsampling), could allow for better rate-distortion control, or other optimizations e.g. of PVQ (perceptual vector quantization) [5], [6].

While DCT transform decorrelates data in a block, there remain other statistical dependencies like homoscedasticity - discussed predicting width (scale parameter) for AC

Improving distribution and flexible quantization for DCT coefficients

Jarek Duda
Jagiellonian University, Golebia 24, 31-007 Krakow, Poland, Email: dudajar@gmail.com

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Evaluation using 48 grayscale 8bit 512x512 images from http://decsai.ugr.es/cvg/CGBase.htm. \textbf{Top left:} log-likelihood evaluation (mean $\log(q(x))$) for 63 AC coefficients of 8x8 DCT-II transform for EPD family $\rho(x) \propto \exp(-(|x-\mu|/\sigma)^\kappa/\kappa)$, shifted to zero for $\kappa = 1/2$ (remaining $\sigma, \mu$ parameters from MLE). Vertical difference can be interpreted as change in bits/pixel for using different $\kappa$: we can see that going from $\kappa = 1$ of Laplace distribution to $\kappa = 1/2$, we get $\approx 0.11$ bits/value mean savings, further individual $\kappa$ optimization gave additional $\approx 0.03$ bits/value mean savings. \textbf{Top right:} rate-distortion comparison for size $N$ uniform quantization (tails go to extremal nodes) on $[-10, 10]$ range of $\kappa = 1/2$, $\sigma = 1$, $\mu = 0$ EPD distribution using this density (blue) or $\kappa = 1$ standard Laplace assumption (orange) - we can see these $\approx 0.1$ bits/value savings from switching to $\kappa = 1/2$, nearly universal for various quantization size $N$. Valuable observation is that quantization into even $N$ is significantly worse - should be avoided, focusing on odd $N$. \textbf{Bottom:} evaluation of various distributions for 8x8 DCT coefficients - perfect agreement would have flat line in zero. Specifically, the values were transformed $y = CDF(x)$ using CDF of assumed distribution, then sorted (empirical distribution function) should ideally give diagonal - which is subtracted. We can see that Laplace has relatively large 0.04 - 0.08 disagreement (orange), it is much smaller for $\kappa = 1/2$ (blue), sometimes a bit further improved for individually optimized $\kappa$. Visually their main imperfection is large jump in the center: corresponding to increased probability of zero value, what can be included in probabilities used for quantization.}
\end{figure}
coefficients based on already decoded coefficients: as linear combination of their absolute values, leads to $\approx 0.5$ bits/value mean savings.

There is also discussed inexpensive automatic approach for quantization especially of such continuous probability distribution functions - by first optimizing quantization density function $q$ describing how choose local density of quantization nodes for asymptotic case of infinite number of region, then use it for finite number of regions. Specifically, $q$ integrates to 1 as density, its inverse CDF (cumulative distribution function) on regular lattice of chosen size $N$ gives the quantization points.

For minimizing distortion for given density $\rho$ - usually MSE (mean squared error) of quantization, like classical Lloyd-Max algorithm [7], [8], here we get $q \propto \rho^{1/3}$, that denser regions should have denser quantization, but only with cube root, e.g. twice denser for 8 times larger density.

However, while such quantization indeed reduces distortion, turns out it also increases entropy by more uniform distribution among quantization regions. Optimizing both rate and distortion, such optimization has lead to nearly uniform quantization (at least for such first considered examples) - with optimized tail handling.

This is work in progress, continuation of author’s revisions of basic approaches for image/video compression [9], [10], for example for context dependent probability distribution models - what is planned to be explored for DCT coefficients in later versions of this article, alongside other expansions.

## II. Exponential Power Distribution (EPD)

For $\kappa > 0$ shape parameter, $\sigma > 0$ scale parameter and $\mu \in \mathbb{R}$ location, probability distribution function (PDF, $\rho_{\kappa\mu\sigma}$) and cumulative distribution function (CDF, $F_{\kappa\mu\sigma}(x)$) of EPD are correspondingly:

$$
\rho_{\kappa\mu\sigma}(x) = \frac{1}{2\sigma} \frac{\kappa^{-1/\kappa}}{\Gamma(1+1/\kappa)} e^{-\frac{\kappa}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{\kappa}} \tag{1}
$$

$$
F_{\kappa\mu\sigma}(x) = \begin{cases} 
\frac{1}{2} \Gamma\left(1, \frac{(x-\mu)/\sigma}{\kappa}\right) & \text{if } x < \mu \\
1 - \frac{1}{2} \Gamma\left(1, \frac{(x-\mu)/\sigma}{\kappa}\right) & \text{if } x \geq \mu 
\end{cases}
$$

where $\Gamma$ is Euler gamma function, $\gamma(a,z) = \Gamma(a,z)/\Gamma(a)$ is regularized incomplete gamma function. Their PDFs for $\kappa = 1/2, 1, 2$ are plotted in [4].

Its (base 2) differential entropy is

$$
H = - \int \rho \log(\rho) = \frac{1}{\kappa \ln(2)} - \log\left(\frac{\kappa}{2\Gamma(1/\kappa)}\right) + \log(\sigma) \tag{2}
$$

for uniform quantization with $\Delta$ lattice size $\Delta Z$, to store such values we need $\approx H - \log(\Delta)$ bits/value.

While ML estimation of $\kappa$ is more difficult, in practice we can often use it as constant - optimized for a given situation, like general AC coefficients, or maybe 63 individual ones for each AC coefficient of 8x8 DCT as considered in Fig. [1]. In many cases like AC coefficients here we can also assume $\mu = 0$, alternatively there can be used MLE approximation as mean value (exact for $\kappa = 2$ Gaussian distribution), or median value (exact for $\kappa = 1$ Laplace distribution), we can also predict it from a context as discussed e.g. in [9], [10].

There remains the main estimation - of width parameter $\sigma$, what turns out quite simple:

$$
\sigma^\kappa = \text{mean } |x - \mu|^\kappa \tag{3}
$$

which can be seen as generalization of the Laplace and Gaussian case, can be naturally turned into context-dependent [9] (e.g. in the next section) or adaptive [4] estimation for nonstationarity.

Here for all 8x8 DCT-II coefficients from 48 grayscale 512x512 images there was calculated log-likelihood for various $\kappa$ - results are presented in Fig. [1]. We can see that $\kappa = 1/2$ fits AC data much better than standard $\kappa = 1$, getting $\approx 0.1$ bits/value reduction. In contrast, DC coefficients have completely different behavior and treatment, here getting optimal $\kappa \approx 2.2$.

## III. Intra-Block Conditional Distributions

DCT transform decorrelates data in block e.g. $8 \times 8$, making additional linear predictions between coefficients inside block rather impractical (experiments suggest $< 0.01$ bpp savings). It still leaves opportunities for inter-block predictions, for example exploiting assumption that DCT in $16 \times 16$ block should also decorrelate well, what allows to predict values in its one of four $8 \times 8$ subblocks based on already decoded three remaining $8 \times 8$ subblocks. It will be explored in further versions of this article, here we focus on additional opportunities inside a single block for AC coefficients.

While DCT allows to exploit correlation of coefficients inside a block, we have also higher statistical dependencies, e.g. between widths of neighboring distributions like homoscedasticity in ARCH-like models. Turns out its exploitation for AC can lead to essential savings: $\approx 0.53$ bits/pixel (total savings for grayscale divided by 64 pixel block, for RGB might be up 3x larger).

Figure [2] shows estimated example of such statistical dependence - with nearly zero correlation but high width dependence, also of $\kappa$ shape parameter if using EPD. It leads to $0.4 - 0.5$ bits/value savings - by estimating width from single already decoded coefficient. Figure [3] shows example of savings and coefficients from automated estimation from already decoded coefficients.

Specifically, for presented standard zigzag order, for each AC $j, k$ position we calculate width of Laplace distribution $\sigma_{jk}$ (centered in $\mu = 0$) as linear combination of absolute values of previous (already decoded) coefficients in zigzag order:

$$
\sigma_{jk} = \beta_{j0} + \beta_{jk12}|DCT_{12}| + \beta_{jk21}|DCT_{21}| + \ldots \tag{4}
$$
While DCT has nearly removed correlations between coefficients, there have remained higher statistical dependencies like homoscedasticity - exploiting of which can bring essential savings for data compression. Here is example of choosing probability distribution for DCT_{13} coefficient, if knowing (already decoded) value of DCT_{12}. The plots were calculated by sorting (DCT_{12}, DCT_{13}) pairs over the first coordinate and focusing on overlapping size 5000 ranges of pairs - mean first coordinate over such range is treated as DCT_{12} in the plots. For second coordinate, in each range there is independently estimated Laplace distribution, we can see that the larger |DCT_{12}|, the larger width \sigma should we choose, with nearly linear dependence. This way we get \approx 0.4 bits/value savings. Bottom: analogously, but estimating more general EPD instead, we can see that additionally \kappa should grow with |DCT_{12}|, increasing savings to \approx 0.5 bits/value. However, trials to essentially improve with varying \kappa were unsuccessful so far (also much more costly), hence there is only used middle \kappa = 1 case.

Where as discussed in [9], [10], \beta coefficients are found with the least squares linear regression to minimize distance to absolute value of predicted coefficient:

\[
\min_{\beta} \text{mean } \left( |DCT_{jk} - \beta_{jk0} - \beta_{jk12} DCT_{12} - \ldots | \right)^2
\]

Surprisingly, the intercept term \beta_{jk0} has nearly negligible effect - we can well estimate \sigma from previous coefficients. Using DC coefficient DCT_{11} for this prediction gives no improvement, however, gradients of DC coefficients of neighboring blocks should be useful, maybe separately treated for vertical and horizontal direction - what is planned to be explored later.

We should avoid negative \beta coefficients as they could lead to problematic negative \sigma. Directly applying such \sigma estimation would require a few dozens multiplications
per value - in practice there are needed approximations reducing it to a few, for example using only a few neighboring already decoded values, maybe also some hidden states like $\sigma$ estimators found for these neighbors (this way containing combination of all previous, can be also states representing already decoded neighboring blocks), using only positive $\beta$ coefficients.

As discussed in [9], we can also apply optimized powers to terms in [4], e.g. for EPD directly estimating $\sigma^p$ instead like variance $\sigma^2$ for Gaussian $\kappa = 2$. There were performed some initial trials, also of predicting $\kappa$ as in bottom of Fig. 2 but without getting essential improvements.

IV. FLEXIBLE DENSITY QUANTIZATION

Having a model of 1D density $\rho : D \rightarrow \mathbb{R}^+$ (integrating to 1, usually $D = \mathbb{R}$) e.g. as Laplace or EPD, there remains crucial question of choosing quantization.

A standard choice is uniform quantization as computationally inexpensive, but it might leave improvement opportunities. On the opposite side there is Lloyd-Max algorithm [11], [8] performing costly mean distortion optimization for a chosen number of regions ($N$). It neglects entropy growth which turns important issue - included in considerations here.

There is discussed approach combining their advantages: for a fixed parametric distribution, we would like to automatize inexpensive process of optimized quantization into flexible number of regions $N$, with control of rate and distortion.

For this purpose, let us introduce quantization density function $q : D \rightarrow \mathbb{R}^+$, also integrating to 1, intuitively defining how dense local quantization should be. We will optimize it accordingly to assumed density $\rho$. Analogously to cumulative distribution function (CDF), let us define:

$$Q(x) = \int_{-\infty}^x q(x)dx \in [0,1]$$  \hspace{1cm} (5)

We can use it to define centers of quantization regions for any number of regions $N$ by taking inverse CDF on some a regular lattice, for example:

$$Q = \{Q^{-1}((i-1/2)/N) : i = 1,\ldots,N\}$$  \hspace{1cm} (6)

For what encoder needs to perform $Q(x)$, e.g. tabled or interpolated in practical realizations, then perform uniform quantization on $[0,1]$. Decoder analogously needs tabled/interpolated $Q^{-1}$ function:

$$\hat{x} = \lfloor NQ(x) \rfloor \quad \hat{x} = Q^{-1}((\hat{x} - 1/2)/N)$$  \hspace{1cm} (7)

Another approach is taking boundaries of quantization regions in the middle between succeeding points of $Q$ - what is used in evaluations.

Optimization of $q$ for a given $\rho$ can be done for $N \rightarrow \infty$ continuous limit, for which we can assume that local distance between quantization nodes in position $x$ is approximately $(Nq(x))^{-1}$.

A. Distortion minimizing quantization

For distortion defined as mean power $p$ of quantization error $|x - \hat{x}|^p$, e.g. $p = 2$ for popular MSE, we can say that mean distortion in position $x$ is proportional to $1/(Nq(x))^p$. Averaging such local distortion over assumed probability distribution $\rho$, we get distortion evaluation:

$$D(q) \equiv D = \int_{x \in D} \rho(x)q(x)^pdx \quad D_N = \frac{D}{N^p}$$  \hspace{1cm} (8)

where $D_N$ is approximation for quantization into $N$ regions. To choose the optimal $q$ we can use calculus of variations (e.g. [11]): to minimize $D$ as in the necessary condition for extremum, the first order correction of $D$ for any (infinitesimal) perturbation $q \rightarrow q + \delta q$ has to be 0, for $\delta q$ being a function integrating to 0 to maintain $\int qdx = 1$.
∀N quantization size (3 different).

Bottom together with such functions providing uniform quantization on largest range (even). We can see that the best ones approach uniform quantization on original distributions: we know their analytical formulas as they are just rescaled family (containing e.g. Laplace and Gaussian distribution),

\[ \int_{\infty}^{\infty} \rho(x) \, dx = 1 \]

\[ \delta_q = \frac{\sigma}{\sqrt{2\pi}} \] for normalization to integrate to 1. For MSE we have \( p = 2 \): quantization density \( q \) should be increased with cube root of density \( \rho \), e.g. twice denser for 8 times larger \( \rho \).

While generally we can find \( q, \rho \) numerically and store in tables for fixed center \( \mu = 0 \) and scale parameter \( \sigma = 1 \) (for shifted and rescaled values), for discussed general EPD family (containing e.g. Laplace and Gaussian distribution), we know their analytical formulas as they are just rescaled original distributions:

\[ \sigma_q = \sigma / \sqrt{p+1} \]

\[ q = \rho_{\kappa \mu \sigma_q} Q = F_{\kappa \mu \sigma_q} \]

for normalization: \( \forall \delta_q \int \delta_q \, dx = 0 \)

\[ 0 = D(q+\delta q) - D = -p \int \frac{\rho(x)}{q(x)} \delta q(x) \, dx \]

It is always zero if \( \rho(x)/(q(x))^{p+1} = \text{const} \). Otherwise, we could increase \( \delta q \) where this fraction is larger, at cost of where it is smaller, getting nonzero variation.

So \( D \) is minimized for \( \rho(x)/(q(x))^{p+1} = \text{const} \), getting:

\[ q(x) = (\rho(x))^{1/(p+1)} / \int (\rho(y))^{1/(p+1)} \, dy \] for normalization to integrate to 1. For MSE we have \( p = 2 \): quantization density \( q \) should be increased with cube root of density \( \rho \), e.g. twice denser for 8 times larger \( \rho \).

While generally we can find \( q, \rho \) numerically and store in tables for fixed center \( \mu = 0 \) and scale parameter \( \sigma = 1 \) (for shifted and rescaled values), for discussed general EPD family (containing e.g. Laplace and Gaussian distribution), we know their analytical formulas as they are just rescaled original distributions:

\[ q = \rho_{\kappa \mu \sigma_q} Q = F_{\kappa \mu \sigma_q} \]

B. Entropy (rate) minimizing quantization

To calculate asymptotic \( N \to \infty \) behavior of entropy (required bits/value rate), probability of quantization region in position \( x \) is asymptotically \( \rho(x)/(Nq(x)) \): requiring \( \lg((Nq(x))/\rho(x)) \) bits.

\[ H_N = \int \rho(x) \lg \left( \frac{Nq(x)}{\rho(x)} \right) \, dx = H + \lg(N) \]

for

\[ H = \int \rho(x) \lg \left( \frac{q(x)}{\rho(x)} \right) \, dx \]

what is minus Kullback-Leibler divergence: gets extremum in \( q = \rho \) (can be obtained with above calculus of variations), but this time maximal number of bits/value - we would like to get far away from it.

The minimal entropy we could get here is usually zero; by quantization which puts practically entire probability into single region, but it makes no sense from practical perspective - optimizing quantization density to minimize entropy alone rather makes no sense.

C. Rate-distortion optimization

Distortion optimization alone indeed reduces it for a given quantization size, however, it happens at cost of increased entropy as it leads to more uniform probability distribution among quantization regions than uniform quantization (should concern also e.g. Lloyd-Max).

Hence, in practice we should optimize distortion and entropy together, what can be done using Lagrange multipliers. We have two constraints here, each gets one multiplier: first for normalization \( \int q \, dx = 1 \) (previously hidden e.g. as \( \text{const} = \rho/(q^{p+1}) \)). Second for fixed entropy or distortion - while minimizing the other, both these cases are mathematically similar.

Finally, we can just use some two multipliers \( \mu, \lambda \), focus on their pairs maintaining normalization \( \int q \, dx = 1 \), getting (entropy, distortion) pairs hopefully being in minimum (not maximum or saddle). For this purpose we can start with the safe: distortion optimization case \( \text{(9)} \) and try to continuously (e.g. numerically) modify it solving ordinary differential equation obtained by treating \( \int q \, dx = 1 \) as implicit equation.

\[ H^1 \text{ quantization error (p = 1): To simplify the solution formula, for p = 1 case let us choose } \mu, \lambda \text{ Lagrange multipliers in the following way:} \]

\[ \mu \frac{\rho}{q^{p+1}} - 2\lambda \rho = 1 \]

\[ \text{satisfying } \int q \, dx = 1 \]

\[ \text{For } \mu, \lambda, \rho, q \geq 0 \text{ we get promising solution:} \]

\[ q = \frac{\rho}{\sqrt{\mu \rho + \lambda^2 \rho^2}} = \frac{\mu \rho}{\sqrt{\mu \rho + \lambda^2 \rho^2 + \lambda \rho}} \]

For \( \lambda = 0 \) we get \( q \propto \rho^{1/2} \) as for distortion minimization. For \( \mu = 0 \) we get the problem of entropy optimization case.

Figure 5 contains such \( (\lambda, \mu) \) pairs satisfying \( \int q \, dx = 1 \), obtained by just optimizing \( \mu \) for succeeding \( \lambda > 0 \) on a lattice. It leads to \( \mu, \lambda \to \infty \) with fixed asymptotic \( \mu/\lambda \). This limit approaches constant \( q \) case of uniform quantization, with additionally optimized handling of tails.
2) **MSE:** $l^2$ quantization error ($p = 2$): For $p = 2$ we get degree 3 polynomial instead of 2, which still has analytical solution, but a bit more complex one - we can perform analogous analysis, what is planned for further versions of this articles.

V. CONCLUSIONS AND FURTHER WORK

While often there are uncritically used assumptions of naturally looking distributions, like Laplace or Gaussian, it might be worth testing also e.g. more general families, like EPD discussed here, or heavy tail like stable distributions appearing e.g. in generalized central limit theorem for addition of i.i.d infinite variance variables. For data compression applications, improvements of likelihood can be directly translated to savings in bits/value. Also, while DCT transform decorrelates coefficients, there remain other statistical dependencies like homoscedasticity - their exploitation is computationally more costly, but as discussed can lead to relatively huge savings.

There was also discussed approach for automatic search of flexible quantization, shifting the problem into finding e.g. continuous quantization density function for $N \to \infty$ continuous limit, and then use it for finite $N$. While it can improve distortion alone, together with entropy optimization it seems to lead to nearly uniform quantization - with additional tail optimization.

This flexible quantization approach definitely needs further work, starting with finishing $p = 2$ case and testing for various distributions. It generally brings a question of practicality also of approaches like Lloyd-Max, what requires deeper exploration.

It also might be valuable to try to expand this work into vector quantization. From classical PVQ perspective, there might be considered deformation to reduce distortion - without it is nearly optimal for Laplace distribution, it can be deformed to optimize for Gaussian distribution with uniform on sphere. we can also use this technique for obtained here deform for $\kappa = 1/2$ EPD distribution.

There is also planned further analysis and testing of context dependent methods from [9], [10] for inexpensive prediction of parameters e.g. $\mu, \sigma$ of $\kappa = 1/2$ EPD distribution, testing statistical dependencies for coefficients inside $8 \times 8$ DCT block and between neighbors, e.g. using observations for larger $16 \times 16$ block treated as four $8 \times 8$ sub-blocks and predicting for one of them from already decoded three.

Further large topic to consider is optimizing transforms, especially color as discussed in [10], maybe also entire DCT-like transform, combination with chroma subsampling for perceptual evaluations.

**REFERENCES**

[1] G. K. Wallace, “The jpeg still picture compression standard,” IEEE transactions on consumer electronics, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[2] J. Minguillon and J. Pujol, “Jpeg standard uniform quantization error modeling with applications to sequential and progressive operation modes,” Journal of Electronic Imaging, vol. 10, no. 2, pp. 475–486, 2001.

[3] P. R. Tadikamalla, “Random sampling from the exponential power distribution,” Journal of the American Statistical Association, vol. 75, no. 371, pp. 683–686, 1980.

[4] J. Duda, “Adaptive exponential power distribution with moving estimator for nonstationary time series,” arXiv preprint arXiv:2003.02149, 2020.

[5] T. Fischer, “A pyramid vector quantizer,” IEEE transactions on information theory, vol. 32, no. 4, pp. 568–583, 1986.

[6] J. Duda, “Improving pyramid vector quantizer with power projection,” arXiv preprint arXiv:1705.05285, 2017.

[7] S. Lloyd, “Least squares quantization in pcm,” IEEE transactions on information theory, vol. 28, no. 2, pp. 129–137, 1982.

[8] J. Max, “Quantizing for minimum distortion,” IRE Transactions on Information Theory, vol. 6, no. 1, pp. 7–12, 1960.

[9] J. Duda, “Parametric context adaptive laplace distribution for multimedia compression,” arXiv preprint arXiv:1906.03238, 2019.

[10] ——, “Exploiting context dependence for image compression with upsampling,” arXiv preprint arXiv:2004.03391, 2020.

[11] V. Arnold, “1., 1978, mathematical methods of classical mechanics,” Graduate Texts in Mathematics, Springer-Verlag, New York, 1982.