COMPRESSIVE SENSING FOR INCOHERENT IMAGING SYSTEMS WITH OPTICAL CONSTRAINTS

Francesco Renna, Miguel R. D. Rodrigues†, Minhua Chen+, Robert Calderbank‡ and Lawrence Carin‡*

* Instituto de Telecomunicações, Departamento de Ciência de Computadores, Universidade do Porto
† Department of Electronic & Electrical Engineering, University College London
+ Department of Statistics, University of Chicago
‡ Department of Electrical Engineering, Duke University

ABSTRACT

We consider the problem of linear projection design for incoherent optical imaging systems. We propose a computationally efficient method to obtain effective measurement kernels that satisfy the physical constraints imposed by an optical system, starting first from arbitrary kernels, including those that satisfy a less demanding power constraint. Performance is measured in terms of mutual information between the source input and the projection measurement, as well as reconstruction error for real world images. A clear improvement in the quality of image reconstructions is shown with respect to both random and adaptive projection designs in the literature.

Index Terms— Compressed sensing, low resolution imaging, projection algorithms

I. INTRODUCTION

Compressive Sensing (CS) is the art of capturing important attributes of a high dimensional signal from a relatively small set of linear projections. It is possible to guarantee fidelity of reconstruction from random projections when the high dimensional signal exhibits low dimensional structure such as sparsity with respect to some dictionary [1], [2]. However, measurements that are aligned with the signal model have recently been shown to provide significant performance improvements over random projections [3], [4]. The models used to represent signals can incorporate structured sparsity [5], [6], manifold descriptions [7] or statistical models [8]–[10]. In the latter case the linear projections should be chosen to align the probability distribution of the signal with that of the measurement noise. In the special case of image processing a Gaussian mixture model (GMM) is known to provide an accurate description of patches extracted from natural images.

Among the metrics that can be used to drive kernel measurement optimization, mutual information between the input source and the projection measurements has been recently proven to provide state-of-the-art results for image reconstruction [10]. In this case, tools from communication theory are imported and adapted to the CS framework to optimize linear projection measurements. However, the constraints taken into account in the optimization process are directly derived from their counterpart in the communication problem, and they do not always reflect the limitations imposed by an actual imaging system.

In this work, we consider incoherent imaging systems in which linear projections of the image are computed directly in the optical domain. Given that the devices implementing the linear projections are optical, specific constraints are imposed on the kind of projections that can be computed. We will present here a procedure to obtain linear projection designs for signal reconstruction that prove effective and that satisfy the constraints imposed by an incoherent optical imaging system.

Existing kernel designs in the literature are compliant with more tractable constraints. For example, sensing matrices are assumed to be drawn from Gaussian or Bernoulli distributions [11], or they are adapted to the signal model, but they only satisfy a power constraint [10]. Other approaches consider the physical limitations introduced by incoherent optical imagining systems [3], [4], but they do not use state-of-the-art methods to adapt the measurement kernels to the input source.

II. SYSTEM MODEL

We consider a compressive imaging system as in Fig. 1, where a naturally illuminated input scene is captured by computing linear projections of the object irradiance distribution, directly in the optical domain [4]. We represent the two-dimensional object irradiance of dimensions $\sqrt{n} \times \sqrt{n}$, where n is an integer, by the column vector $\mathbf{x} \in \mathbb{R}^n$. We assume that $\mathbf{x}$ can be described as a random vector whose probability density function (pdf) $p(\mathbf{x})$, is known, or can be learned from a training data set. The source irradiance is collected over a predefined time interval by photon detectors,

The work of F. Renna and M. R. D. Rodrigues has been supported by Fundação para a Ciência e a Tecnologia through the research project CMU-PT/SIA/0026/2009.

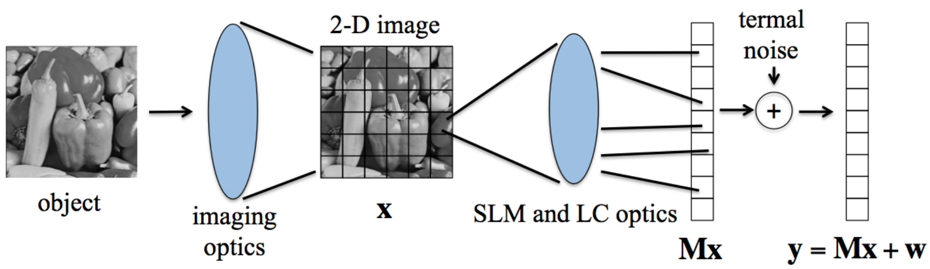

Fig. 1. Schematic representation of the incoherent optical imaging system. Linear projections are computed directly in the optical domain.

and the elements of $\mathbf{x}$ represent the number of photons corresponding to a particular region of the two dimensional image plane. Linear projections performed in the optical domain are modeled by multiplying $\mathbf{x}$ by the measurement matrix (kernel) $\mathbf{M} \in \mathbb{R}^{m \times n}$, in which we consider in general[1] $m \ll n$. The (small) set of noisy measurements from which the signal $\mathbf{x}$ will be reconstructed is represented by the vector

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}, \tag{1}$$

in which $\mathbf{w}$ is a Gaussian vector with zero mean and covariance matrix $\mathbf{\Sigma_w} = \mathbb{E}\left[\mathbf{w}\mathbf{w}^{\mathrm{T}}\right]$, accounting for the termal noise introduced after photo-current conversion[2].

The linear projections performed by the matrix $\mathbf{M}$ can be seen to arrange the photons emitted by the source among the small number of measurements, contained in the vector $\mathbf{y}$, and due to conservation of energy, the total number of photons is given (and finite). Therefore, the projection kernels that are implementable in the optical domain must satisfy the following constraints,

$$0 \leq m_{ij} \leq a \ \forall i, j \tag{2a}$$

$$\sum_i m_{ij} \leq a \ \forall j, \tag{2b}$$

where m_{ij} is the i-th row, j-th column entry of $\mathbf{M}$. Note that the constant $a > 0$ permits a general mathematical formulation of the problem. However, we will include numerical results for the case $a = 1$, as this is the only value reflecting a physical implementation of the system.

III. KERNEL DESIGN WITH OPTICAL CONSTRAINTS

III-A. Minimum Frobenius norm kernel design

The kernel design problem subject to the constraints in (2) is difficult to solve for various performance metrics, such as mutual information and minimum mean square error (MMSE). Therefore, we propose here a minimum Frobenius norm method to obtain effective measurement kernels that are compliant with the physical constraints. Namely, the proposed design can be described by the following two steps procedure.

1) Identify an effective measurement matrix $\mathbf{M}_0$ which may not satisfy the constraints in (2), but that is shown to provide good performance in terms of a predefined metric.
2) Compute the matrix $\mathbf{M}_{\mathrm{md}}$ that minimizes the Frobenius norm distance from $\mathbf{M}_0$ and satisfies the optical constraints.

A complete characterization of the proposed projection design is obtained by solving the optimization problem

$$\begin{aligned} \underset{\mathbf{M}}{\text{minimize}} \quad & \|\mathbf{M}_0 - \mathbf{M}\|_{\mathrm{F}} \\ \text{subject to} \quad & 0 \leq m_{ij} \leq a, \ \forall i, j, \\ & \sum_i m_{ij} = a, \ \forall j, \end{aligned} \tag{3}$$

where the symbol $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix. On denoting the entries of $\mathbf{M}_0$ by $m_{0,ij}$, the solution of the problem (3) is described in the following theorem.

Theorem 1: The solution of the minimum Frobenius norm problem in (3) can be characterized as follows,

$$m_{\mathrm{md},ij} = \left[m_{0,ij} - \frac{\sum_{i \in \mathcal{P}_j} m_{0,ij} - a}{|\mathcal{P}_j|} \right]^+ \tag{4}$$

with $[x]^+ = \max\{0, x\}$, where $\mathcal{P}_j$ denotes the set of indexes i for which the corresponding $m_{\mathrm{md},ij}$ in the optimal solution is greater than zero.

Proof: The proof follows directly from the evaluation of the Karush-Kuhn-Tucker (KKT) conditions [12] associated to the optimization problem in (3). ■

Remark 1: The characterization of the minimum Frobenius norm solution in Theorem 1 is implicit, as the term $\mathcal{P}_j$ depends on the values of the optimal $m_{\mathrm{md},ij}$'s. However, an efficient implementation, which we leverage in Section IV, can be devised by considering a geometrical interpretation

[1] Each row of the projection matrix $\mathbf{M}$ is physically implemented using a spatial light modulator (SLM), followed by light-collection (LC) optics that directs the light to photon detectors.

[2] For the sake of mathematical tractability of the model, we neglect the effect of the signal shot noise (Poisson noise) and of the dark current associated to photon detectors.

of the problem. Namely, each column in $\mathbf{M}_{\text{md}}$ is a vector in $\mathbb{R}^m$ that is obtained by projecting the corresponding column of $\mathbf{M}_0$ onto a rescaled version of the canonical simplex of $\mathbb{R}^m$. The computation of such a projection is a well studied problem and fast solutions have been proposed. Among the results in the literature, we implemented the iterative algorithm in [13], as it is proved to converge to the correct value of $\mathcal{P}_j$ in at most m iterations.

III-B. Mutual information based kernel design

We now consider mutual information based kernel design, that aims at maximizing the mutual information between the input source and the projection measurements, $\mathbb{I}(\mathbf{x};\mathbf{y})$. This quantity has been recently proved to be an effective proxy to adaptively design measurement projections for CS [10]. The reason lies in the fact that $\mathbb{I}(\mathbf{x};\mathbf{y})$ represents the amount of information over $\mathbf{x}$ that can be extracted from the observation of the random (measurement) vector $\mathbf{y}$. However, the mutual information maximization problem has been solved for the case in which only a trace constraint is imposed on the matrix $\mathbf{M}\mathbf{M}^{\mathrm{T}}$ [10]. Hence, we compute $\mathbf{M}_0$ as the solution of

$$\begin{aligned} &\underset{\mathbf{M}}{\text{maximize}} && \mathbb{I}(\mathbf{x};\mathbf{y}) \\ &\text{subject to} && \operatorname{tr}(\mathbf{M}\mathbf{M}^{\mathrm{T}}) \le b. \end{aligned} \tag{5}$$

Different values of the trace constraint b will correspond to different $\mathbf{M}_0$'s, and, consequently, to different minimum Frobenius norm kernel designs. Therefore, we decide to choose b as the smallest value for which the problem in (5) is a relaxed version of the mutual information maximization problem with constraints (2). In particular, it is easy to verify that if $b = na^2$, then the set (2) is contained in the feasible set of (5). In fact, the set of matrices verifying $\operatorname{tr}(\mathbf{M}\mathbf{M}^{\mathrm{T}}) \le na^2$ contains the set of matrices having ℓ_2 norm over each column less or equal than a^2. Moreover, as $0 \le m_{ij} \le a,\ \forall i,j$, then $\sum_i m_{ij} \le a \Rightarrow \sum_i m_{ij}^2 \le a^2,\ \forall j$.

Given this characterization of $\mathbf{M}_0$, the minimum Frobenius norm approach is also justified by the regularity properties of mutual information and, namely, by the following result.

Theorem 2: For all matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ such that $\operatorname{tr}(\mathbf{M}_1\mathbf{M}_1^{\mathrm{T}}) \le na^2$ and $\operatorname{tr}(\mathbf{M}_2\mathbf{M}_2^{\mathrm{T}}) \le na^2$,

$$|f(\mathbf{M}_1) - f(\mathbf{M}_2)| \le L\ \|\mathbf{M}_1 - \mathbf{M}_2\|_{\mathrm{F}}, \tag{6}$$

where $f(\mathbf{M}) = \mathbb{I}(\mathbf{x};\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w})$, $\boldsymbol{\Sigma}_{\mathbf{x}}$ is the input covariance matrix and $L = a\ \operatorname{tr}(\boldsymbol{\Sigma}_{\mathbf{x}})\sqrt{n\operatorname{tr}(\boldsymbol{\Sigma}_{\mathbf{w}}^{-2})}$.

Proof: The proof is not reported here for reasons of space, and it is based on the characterization of the Jacobian matrix of the function f, as given in [14]. ■

Note that Theorem 2 states that the function $f(\mathbf{M})$ is L-Lipschitz over the set of matrices $\mathbf{M}$ verifying the trace constraint $\operatorname{tr}(\mathbf{M}\mathbf{M}^{\mathrm{T}}) \le na^2$. We conclude that for all $\mathbf{M}$ satisfying the optical constraints in (2),

$$|f(\mathbf{M}_0) - f(\mathbf{M})| \le L\ \|\mathbf{M}_0 - \mathbf{M}\|_{\mathrm{F}}. \tag{7}$$

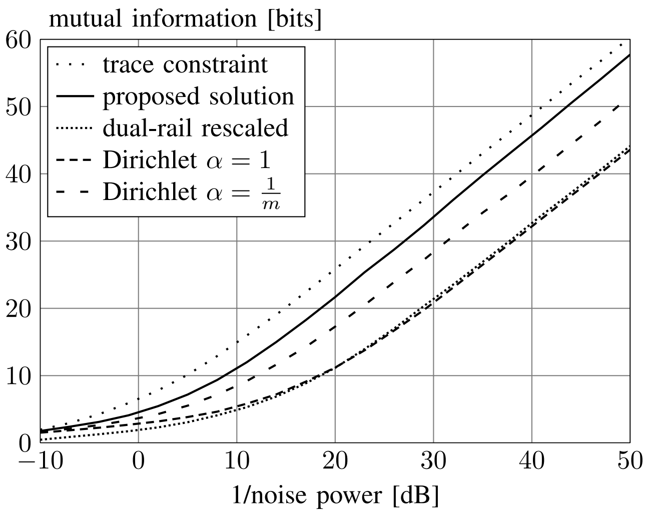

Fig. 2. Mutual information vs. inverse of the noise power. $n = 64, m = 10$. GMM input source (trained from Berkley Segmentation Dataset). Comparison of mutual information values obtained with different projection designs. Optimal trace constrained, proposed solution, dual-rail rescaled implementation [4], random Dirichlet matrices.

Therefore, this justifies our approach because minimizing the Frobenius distance $\|\mathbf{M}_0 - \mathbf{M}\|_{\mathrm{F}}$ also minimizes an upper bound to the mutual information loss with respect to the solution of the relaxed optimization problem with trace constraint. In other terms, our approach consists in this case to solving a relaxed version of the mutual information maximization problem and then finding a feasible matrix that approaches performance close to that guaranteed by the optimal solution of the relaxed problem.

IV. NUMERICAL RESULTS

We present numerical results that compare the proposed measurement design with both random and adaptive measurement approaches in the literature. In Fig. 2 are reported the values of mutual information (versus the inverse of the noise power) obtained with a synthetic GMM input source. In particular, the input pdf $p(\mathbf{x})$ is a 20-component GMM that is trained on 100,000 patches of size 8×8 pixels, randomly extracted from 500 natural images in the Berkeley Segmentation Dataset[3] according with the dictionary learning method described in [9].

The matrix $\mathbf{M}_0$ is the solution of the problem in (5) with $a = 1$, and is computed via the method described in [10]. Then, the proposed solution $\mathbf{M}_{\text{md}}$ is compared with the following designs satisfying the optical constraints. We consider the dual-rail implementation $\mathbf{M}_{\text{dr}}$ described

[3]http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html

in [4], for which two complementary arms are devoted to represent separately the positive and negative values of the projection matrix. As a consequence, this approach results in doubling the noise covariance. Moreover, in order to verify the total photon counting constraint, the measurement matrix is obtained as $\mathbf{M}_{\mathrm{dr}} = \mathbf{M}_0/K$, in which the normalization constant is defined as $K = \max_j \sum_i |m_{0,ij}|$.

Results obtained with random projection matrices are also shown: in particular, we have considered the case in which the columns of $\mathbf{M}$ are independent identically distributed (i.i.d.) Dirichlet random vectors [15], that is when the vectors $[m_{1j}, \ldots, m_{mj}]^{\mathrm{T}}$ are independently drawn from the distribution

$$p(x_1, \ldots, x_m; \alpha_1, \ldots, \alpha_m) = \frac{1}{B(\alpha_1, \ldots, \alpha_m)} \prod_{i=1}^{m} x_i^{\alpha_i - 1} \tag{8}$$

in which $x_1, \ldots, x_{m-1} \geq 0$, $x_m = 1 - \sum_{i=1}^{m-1} x_i$ and $B(\alpha_1, \ldots, \alpha_m)$ is the multinomial Beta function [16]. Note that, for all choices of the concentration parameters α_i, the corresponding random measurement matrices satisfy the optical constraints with $a = 1$. In particular, we have considered the two cases for $\alpha_i = \alpha = 1$ and $\alpha_i = \alpha = 1/m$. From the results in Fig. 2, it is possible to observe that the proposed method offers a gain of approximately $10\,\mathrm{dB}$ in the medium/low-noise regime with respect to the dual-rail rescaled implementation and of approximately $2.5\,\mathrm{dB}$ with respect to Dirichlet measurements with concentration parameter $\alpha = 1/m$. This choice of α guarantees better performance than $\alpha = 1$ in light of the fact that, on average, the square values of the entries in $\mathbf{M}$ are decreasing functions of α. In fact, it is straightforward to show that, when the column of the measurement matrix are sampled from a Dirichlet distribution with concentration parameter α, then $\mathbb{E}\left[\|\mathbf{M}\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[\mathrm{tr}(\mathbf{M}\mathbf{M}^{\mathrm{T}})\right] = n\frac{\alpha+1}{m\alpha+1}$. Hence, a smaller value of α guarantees higher squared amplitude of the measurements, and eventually a higher output SNR.

In the following, we evaluate reconstruction fidelity in terms of the peak signal-to-noise ratio (PSNR) and show experimental results on the performance of the proposed measurement design when applied to image reconstruction. We consider the widely analyzed 256×256 image "pepper" and divide it into 1024 non-overlapping patches of dimension 8×8. The patches are assumed to be statistically described with the GMM signal model learned on the (independent) Berkeley database, and are reconstructed via the conditional mean estimator $\hat{\mathbf{x}}(\mathbf{y}) = \mathbb{E}\left[\mathbf{x}|\mathbf{y}\right]$ on the basis of the measurements obtained with the different projection matrices described in previous paragraphs.

The PSNR values obtained with the previously described projection matrices with various noise power levels are shown in Fig. 3. The proposed method provides an advantage of more than $2\,\mathrm{dB}$ of PSNR with respect to the other methods satisfying the optical constraints throughout a wide range

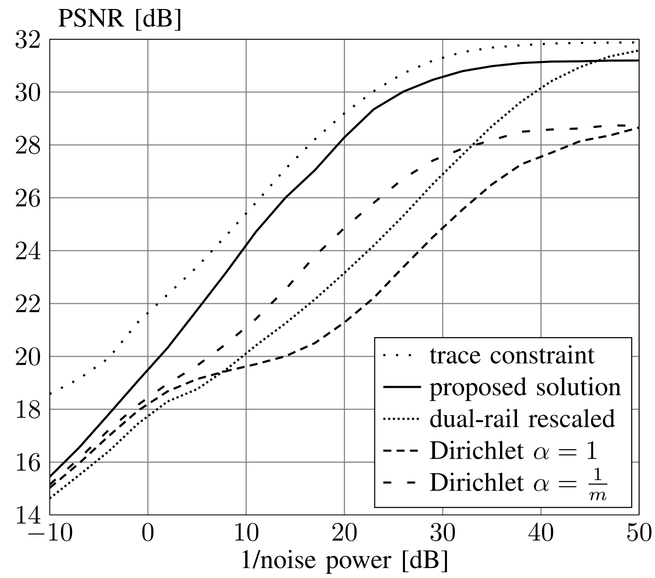

Fig. 3. Image "pepper". PSNR vs inverse of the noise power. Number of projections, $m = 20$. Comparison of the performance obtained with different projection designs. Optimal trace constrained, proposed solution, dual-rail rescaled implementation [4] , random Dirichlet matrices.

of noise power levels. Only when the impact of noise is negligible, the minimum Frobenius norm method is outperformed by the dual-rail implementation $\mathbf{M}_{\mathrm{dr}}$. This result can be explained by noting that, in the low-noise regime, rescaling the measurement matrix $\mathbf{M}_0$ induces mainly a noise amplification effect at the reconstruction stage and hence a shift to the left of the PSNR curve. However, since the PSNR saturates when the noise power is approaching zero, this shift does not affect the overall performance.

V. CONCLUSIONS

A simple design approach for linear projection measurements in CS has been studied. The method is applicable to incoherent optical imaging systems and aims at enabling reliable image reconstruction from low dimensional projection measurements. The proposed method represents a way to effectively adapt a given measurement design to the particular constraints imposed by the physical limitations of the incoherent optical system. It is based on the evaluation of projections onto the canonical simplex of $\mathbb{R}^m$, that can be efficiently computed by using fast algorithms in the literature. This approach is shown to manifest encouraging gains in terms of mutual information between the input source and the measurements, both with respect to random projections and adaptive kernel designs based on dual rail-implementations. Finally, test data on a natural image show the advantage obtained with this method in terms of reconstruction PSNR, for a wide range of noise levels.

VI. REFERENCES

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery for incomplete and inaccurate measurements," *Commun. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, Mar. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] A. Ashok, P. Baheti, and M. Neifeld, "Compressive imaging system design using task-specific information," *Applied Optics*, vol. 47, no. 25, pp. 4457–4471, 2008.

[4] P. Baheti and M. Neifeld, "Recognition using information-optimal adaptive feature-specific imaging," *J. Opt. Soc. Amer. A*, vol. 26, no. 4, pp. 1055–1070, 2009.

[5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[6] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Sig. Proc.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.

[7] R. G. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.

[8] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Sig. Proc.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[9] M. Chen, J. Silva, J. Paisley, C. Wang, D. Dunson, and L. Carin, "Compressive sensing of manifolds using a nonparametric mixture of factor analyzers: algorithm and performance bounds," *IEEE Trans. Sig. Proc.*, vol. 58, no. 12, pp. 6140–6155, Dec. 2010.

[10] W. R. Carson, M. R. D. Rodrigues, M. Chen, L. Carin, and R. Calderbank, "Communication-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[11] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, dec. 2005.

[12] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[13] C. Michelot, "A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbb{R}^n$," *J. Opt. Theory Appl.*, vol. 50, no. 1, pp. 195–200, Jan. 1986.

[14] M. Payaro and D. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.

[15] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions. Volume 1: Models and Applications*. New York: Wiley, 2000.

[16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York, NY, USA: Dover Publications, INC., 1965.