REGULARIZATION METHODS IN THE PROBLEM ON THE RECONSTRUCTION OF OPTICAL SIGNAL. REGULARIZATION METHODS BASED ON RIESZ LEMMA

S.M. Chernyavskii

A.N. Tupolev State Technical University, Kazan'
Received July 2, 1996

In this paper I propose the regularization methods in application to solution of inverse problems of optics, namely, the method based on Riesz lemma in the problem on reconstruction of incoherent source from its noisy image and the amplitude modulation method in the problem on complex signal reconstruction based on its autocorrelation function.

Reconstruction of the incoherent source from its noisy image. The intensity distribution of radiation from a source of monochromatic light $I_0(\mathbf{x}_0)$, $\mathbf{x}_0 = (x_0, y_0) \in E_0$ and the intensity distribution $I(\mathbf{x})$, $\mathbf{x} = (x, y) \in E$ of its noisy image are related by the superposition integral[1]

$$I(\mathbf{x}) = \iint\limits_{E_0} h(\mathbf{x},\mathbf{x}_0)\, I_0(\mathbf{x}_0)\, \mathrm{d}\mathbf{x}_0 + n(\mathbf{x}), \quad \mathbf{x} \in E \subset E_0. \quad (1)$$

Here the kernel $h(\mathbf{x}, \mathbf{x}_0)$ is a given function and in the case of nonisoplanatic function of point scattering, the function $n(\mathbf{x})$ describes the image noise.

The problem on calculation of $I_0(\mathbf{x}_0)$ from Eq. (1) is an ill-posed problem.[2] One of the regularization methods used to solve it consists in assuming certain properties of the solution sought *a priori* thus narrowing the set of solutions. In addition to the property of nonnegative character, the solution is assumed to be smooth and certain limitations are imposed on this solution, which can be statistical. Such limitations can be taken into account by selecting the solution space. The Hilbert space is frequently used in this case. The norm value of the Hilbert space serves as a regularizing factor.

Since the noise $n(\mathbf{x})$ is also unknown as $I_0(\mathbf{x}_0)$, it is natural to take for the solution of Eq. (1) a pair of functions $(I_0(\mathbf{x}_0), n(\mathbf{x}))$ with the given limitations:

$$I_0(\mathbf{x}_0) \in U \subset H, \quad n(\mathbf{x}) \in V \subset H_1, \quad (2)$$

where the Hilbert spaces H and H_1 as well as the sets U and V (being normally convex) determine the type of regularization of Eq. (1).

Analysis of the known regularization methods,[2,3] namely, the Tikhonov regularization using a regularizing functional, the method of maximum likelihood, the method of maximum entropy, has shown that the Eq. (1) can be solved much simpler if this equation is described in the terms of the scalar product of the Hilbert space.

At a fixed value of $\mathbf{x} \in E$ the function $h(\mathbf{x}, \mathbf{x}_0)$ of the variable $\mathbf{x}_0 \in E_0$ is given by the functional on H, and its type is determined by the integral term in Eq. (1). We have a family of the functionals $h_{\mathbf{x}}(I_0)$ on H, depending on the parameter $\mathbf{x}$. According to the Riesz lemma,[4] any continuous functional $h(I)$ on H with the norm $\|h\|$ has its single representation in the form of a scalar product (φ_h, I) of this space, in this case, $\|h\|^2 = (\varphi_h, \varphi_h)$. The proof of this lemma is constructive, it gives an explicit form of an element $\varphi_h = h(\varphi_1)(\varphi_1, \varphi_1)^{-2}\varphi_1$, where φ_1 belongs to an orthogonal complementary minor of the subspace of the functional h zeros: $h(I) = 0$. The element φ_h can be found as the solution of the problem

$$(\varphi, \varphi) \to \min \quad \text{at} \quad h(\varphi) = 1. \quad (3)$$

The Riesz lemma enables one to represent the integral term in Eq. (1) in the form $h_{\mathbf{x}}(I_0) = (\varphi_{\mathbf{x}}, I_0)_H$, where $\varphi_{\mathbf{x}}$ is the solution of the problem (3) at $h_{\mathbf{x}}(\varphi) = 1$.

The noise $n(\mathbf{x})$ can be also considered as the value of a certain series of functionals on H_1, depending on the parameter $\mathbf{x} \in E$ and all the considerations can be repeated for it. But one can also follow different way. We shall restrict our consideration to the case when the set E is finite: $E = \{\mathbf{x}_k,\ k = 1,\ldots,K\}$. The noise value $n(\mathbf{x}_k)$ can be considered as the value of Euclidean scalar product of the vector $n \in R^K$ and the basis vector $e_k = (0, 0,\ldots,1,\ldots,0) \in R^K$, whose kth coordinate equals unity. For R^K another scalar product can be assigned, converting it to the space H_1: $(n_1, n_2)_H = (Bn_1, n_2)$, where B is the positive definite matrix. According to the Riesz lemma there exists a single element $\phi_k \in H_1$, such that $n(\mathbf{x}_k) = (e_k, n) = (B\phi_k, n)$. It is evident that $\phi_k = B^{-1}e_k$. The problem (1)–(2) is now written as

$$I_k = (\varphi_k, I_0) + (B\phi_k, n),\ I_0 \in U,\ n \in V,\ k = 1, \ldots, K, \quad (4)$$

and can be considered as the problem on seeking of the functional, which is determined by the pair

0235-6880/96/11 977-04 $02.00

$(I_0, n) \in U \times V \subset H \times H_1$ and for the given elements $(\varphi_k, \phi_k) \subset H \times H_1$ the functional takes the values I_k. This is the so-called finite-dimensional problem of moments (FDPM).[5] The necessary and sufficient condition for the FDPM to be solved reduces to the solution of the inequality

$$\min_{\lambda} \sup_{(I_0,n)\in U\times V} [(\Phi(\lambda),I_0) + (B\ \Psi(\lambda), n)] = \gamma \geq 1 \quad (5)$$

at any $\lambda = (\lambda_1,\ldots,\lambda_k) \in R^K$ satisfying the condition $\sum_{k=1}^{k} \lambda_k I_k = 1$, where $\Phi(\lambda)$ and $\Psi(\lambda)$ are the linear combinations of the functions $\{\varphi_k\}$ and $\{\phi_k\}$ with the coefficients $\{\lambda_k\}$. If λ^0 is the point of minimum of the problem (5), then among the solutions of FDPM one can distinguish the solution (I^0, n^0) with an extremum property:

$$(\text{t}\ (\lambda^0), I^0) + (B\Psi(\lambda^0), I^0) =$$

$$= \sup_{(I_0,n)\in U\times V} [(\text{t}\ (\lambda^0), I^0) + (B\Psi(\lambda^0), n)]. \quad (6)$$

Examples of regularizing spaces H, H_1 and sets U, V.

1. We shall seek the solution to the FDPM in the class of continuous functions. Then it is reasonable to assume that H is the Sobolev space $W_2^1(E_0)$ with the scalar product

$$(\varphi_1, \varphi_2) = \iint_{E_0} (\text{grad}\ \varphi_1 \cdot \text{grad}\ \varphi_2) + \mu\varphi_1\varphi_2)\text{d}\mathbf{x}_0, \quad \mu \geq 0.$$

The elements φ_k of the problem (3) satisfy the necessary condition of extremum, which results in the equation

$$-\Delta\varphi + \mu\varphi = h(\mathbf{x}_k, \mathbf{x}_0) \quad (7)$$

with the boundary condition $\partial\varphi/\partial n = 0$ to $\text{d}E_0$, where Δ is the Laplacian operator. If the limitation to the FDPM solution is taken as

$$\{(I_0, n)\text{: } I_0 = I_{\text{med}} + u, \ \ I_{\text{med}} > 0,$$

$$f(u, n) = (u, u) + \alpha(Bn, n) \leq l^2, \quad \alpha > 0\},$$

we obtain the solution, regularized by Tikhonov method using the functional $f(u, n)$, which has an extremum property (6) and the structure $u^0 = \Phi((l/\gamma)\lambda_0)$, $n^0 = \Psi((l/\gamma)\lambda^0)$. Thus, the pair $(I_0, n) = (I_{\text{med}} + \Phi(\lambda),\Psi(\lambda))$ is the solution of the system (4) and the system itself determines the vector of coefficients λ.

The usefulness of the approach proposed is illustrated with a one-dimensional case. Assuming that $x_0 = t$ and $\mu = \alpha^2$, the problem (7) takes the form:

$$-\varphi'' + \alpha^2\varphi = h(t), \quad \varphi'(a) = \varphi'(B) = 0. \quad (8)$$

The problem (8) has a unique solution in quadratures. So, only the value λ should yet be calculated from the set of linear algebraic equations to finally determine the solution (I_0, n). For a comparison, if the solution regularized by Tikhonov method is to be sought by minimizing the regularizing functional, it should be determined from the integro-differential equation.

From the equation $h(I_0) = (\varphi_h, I_0)$, following from the Riesz lemma, it is evident that the transform $h \to \varphi_h$ is linear and continuous. In the first example it coincided with the inverse operator of the positive definite operator $A\varphi$ of the problem (7). This means that the Hilbert space H in the first example coincided with the energy space[6] of the operator A: $(\varphi, \varphi)_H = (A\varphi, \varphi)_L$, where $L = L_2(E_0)$ is the space of the functions with the summable squares on E_0. Hence it follows that the regularization of the problem (1) considered here amounts to the determination of I_0 in the class of generalized solutions of a certain positive definite operator with a given or minimal energy norm of this operator.

2. Supplementary information on the source $I_0(\mathbf{x}_0)$ is often of statistical character. The function $I_0(\mathbf{x}_0)$ is considered as realization of a certain random process with a preset mathematical expectation $MI_0(\mathbf{x}_0)$ and the correlation function $R_I(\mathbf{x}_0', \mathbf{x}_0'')$. Then one can take the energy space as the space H corresponding to an operator inverse to the integral operator:

$$Ah = \iint_{E_0} R_I(\mathbf{x}', \mathbf{x}'')h(\mathbf{x}'')\text{d}\mathbf{x}'' + \mu h(\mathbf{x}'), \ \ \mu > 0,$$

with the scalar product $(\varphi, I_0)_H = (A^{-1}\varphi, I_0)_L$.

If the noise is also given by the mathematical expectation M_{nk} and the correlation matrix R_K, then the scalar product in H_1 can be given, as earlier, by the matrix $B = R_K^{-1}$.

In the final analysis the problem (1) amounts to FDPM

$$c_k = (A^{-1}\varphi_k, \overset{\circ}{I}_0) + (R_K^{-1}\ \phi_k,n), \quad k = 1,\ldots,K, \quad (9)$$

where $c_k = I_k - h_k(MI_0)$; $\varphi_k = Ah_k$; $\overset{\circ}{I}_0$ is the deviation of I_0 from MI_0.

If the source I_0 and the noise are independent normal random processes, then the square of the pair (I_0, n) norm on $H \times H_1$ is determined by the equality

$$|(\overset{\circ}{I}_0, n)|^2 = (A^{-1}\ \overset{\circ}{I}_0, \overset{\circ}{I}_0) + (R_K^{-1}\ n, n) \quad (10)$$

and is an analog of a quadratic form in the exponent, determining *a posteriori* density of finite-dimensional distributions in the Bayes formula.[2] According to the Bayes approach, the best estimate of the source I_0 will correspond to the solution of equation (1), where the value of Eq. (10) is minimal. Thus, the problem (1) is reduced to seeking a solution to FDPM with a minimal norm. Reasonings in example **1** show that the solution (I_0, n) is a linear combination of pairs (φ_k, ϕ_k) whose coefficients are defined by the equalities (9).

3. Let the space $H = L_2(E_0)$. Then the integral term in Eq. (1) sets the scalar product in $L_2(E_0)$ at once without any transforms of the elements h_k. The sets U and V are determined by the expressions

$$U(l) = \{I_0 : I_0 = 0.5I_{\max} + 0.5I_{\max}u,\ |u| \le 1,$$
$$\iint_{E_0} |u(\mathbf{x}_0)|^p \mathrm{d}\mathbf{x}_0 \le l^p\}, \qquad (11)$$

$$V = \{n : (Bn,\ n) \le \delta^2\}, \qquad (12)$$

where $I_{\max} = I_{\max}(\mathbf{x}_0)$ is the estimate of the source intensity maximum at the point $\mathbf{x}_0$; the limitation $|u| \le 1$ provides the nonnegative character of $I_0(\mathbf{x}_0)$; δ is the assigned value, characterizing the noise intensity.

Let $l_{\min}$ be the least value of l, at which FDPM is solvable at $(u, n) \in U(l)\times V$. The solution of FDPM on $U(l_{\min})\times V$ is regularized. The reconstructed source $I_0(\mathbf{x}_0)$ at limitations (11) and (12) has the feature that among negative functions on E_0 it has the least deviation from $I_{\max}/2$ in the metrics, determined by the second inequality in Eq. (11). This type of regularization is similar to the regularization using the method of maximal entropy.[2]

Now we write the condition for the FDPM (5) to be solved in equivalent form

$$\max_{(B\lambda,\lambda)\ll 1} \left\{ \sum_{k=1}^{K} c_k\lambda_k - \max_{u\in U(l)} (0.5I_{\max} \sum_{k=1}^{K} h_k\lambda_k,\ u) \right\} = \delta_0 \ll \delta,$$

where $c_k = I_k - (h_k,\ 0.5I_{\max})$.

The value $I_{\min}$ is found from the condition $\delta_0 = \delta$. In this connection it is useful to note that at $l = 0$ the value $\delta_0 = (Bc,\ c)^{1/2}$ is maximum. Therefore the value of δ must be less than the above value. By varying $I_{\max}$ we can gain the fulfillment of the conditions $\delta_0 < \delta$ at $l = +\infty$, i.e., when the second limitation in Eq. (11) is lacking. Introduction of the second limitation in Eq. (11) suggested that with the decrease of l the set $U(I)$ narrows and δ_0 increases, therefore $l_{\min}$ exists at which $\delta_0 = \delta$. One can show that the function $u \in U(l_{\min})$ having an extremum property (6) is a solution of the FDPM, which continuously depends on c_k, i.e., on the initial data on the norm in $L_p(E_0)$.

Reconstruction of the complex function Based on its autocorrelation function. This is one of the frequently occurring inverse problems in optics: the complex function $G(\xi', \eta')$, $(\xi', \eta') \in \Omega$ is found from the equation

$$\iint_{\Omega} G(\xi + \xi',\ \eta + \eta')\, G^*(\xi',\ \eta')\, \mathrm{d}\xi'\mathrm{d}\eta' = H(\xi,\ \eta), \qquad (13)$$

where $H(\xi, \eta)$ is known and has the complex conjugate symmetry.

The question on uniqueness of the solution to Eq. (13) has been studied completely. If $G(\xi', \eta')$ is the solution of Eq. (1) then, except for rare cases, its set of solutions $\{G(\xi', \eta')\exp(\varphi),\ G^*(-\xi', -\eta')\exp(-\varphi),\ \varphi - \text{const}\}$ has unessential manifold.

Equation (1) has square nonlinearity and is solved by iteration methods. Such important problems as the choice of initial approximation, regularization of the problem on seeking of solution are mainly solved using simulations. There is a case when these questions can be answered on the basis of the developed theory. Here, we are dealing with a holographic approach,[7] when $\Omega = W_1 + \omega$, in this case the areas W_1 and ω are well fairly spaced. If the function G is set on ω, Eq. (13) becomes linear relative to G on W_1 given a corresponding selection of the shift vector (ξ, n).

This example shows also that the aperture configuration Ω affects the method of the equation solution. Below we develop this approach. Such configurations apertures were obtained, which enable one to construct new methods for solving Eq. (13).

Let the original aperture Ω be the square with the side $2a$. We divide it into N vertical bands Ω_s of width $\Delta = 2a/N$ (Fig. 1). Narrowing of the function $G(\xi', \eta')$ on Ω_s is denoted by $G_s(\xi', \eta')$ and the set of the shift vectors is given as follows

$$E_0 = \{(\xi_k,\ \eta) : \xi_k = -k\Delta,\ \ k = 0, \ldots, N-1 \text{ and } |\eta| \le 2a\}.$$

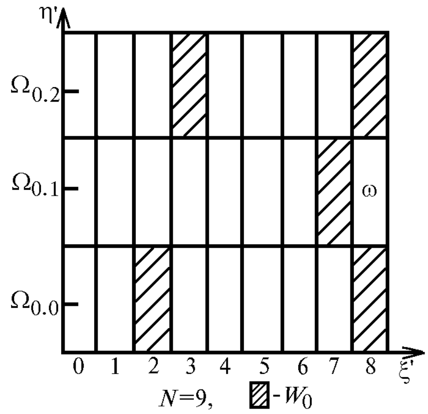

FIG. 1.

Equation (13) with respect to the function $G(\xi', \eta')$ reduces to the set of equations relative to $G_s(\xi', \eta')$:

$$H(-k\Delta, \eta) = \sum_{p=0}^{N-1-k} \int_0^{\Delta} \mathrm{d}\xi' \int_0^{2a} G_p(\Delta p + \xi',\ \eta + \eta') \times$$
$$\times\, G^*_{p+k}(\Delta(p + k) + \xi',\ \eta')\mathrm{d}\eta', \quad k = 0, \ldots, N-1. \qquad (14)$$

If the external integral in Eq. (14) is calculated approximately using the method of mean rectangles and the following designations are introduced, then

$$G_p(\eta') = G_p(\Delta(p + 0.5),\ \eta') \text{ and } H(-k\Delta,\ \eta) = H_k(\eta),$$

the set of equations (14) is transformed to the form:

$$\int_0^{2a} G_0(\eta + \eta')G^*_{N-1}(\eta')\mathrm{d}\eta' = H_{N-1}(\eta)/\Delta, \quad (k = 0); \qquad (15)$$

$$\int_0^{2a} G_0(\eta+\eta')G^*_{N-2}(\eta')d\eta' + \int_0^{2a} G_1(\eta+\eta')G^*_{N-1}(\eta')d\eta' =$$

$$= H_{N-2}(\eta)/\Delta, \quad (k = 1); \qquad (16)$$

$$\int_0^{2a} G_0(\eta+\eta')G^*_{N-1-k}(\eta')d\eta' +$$

$$+ \int_0^{2a} G_k(\eta+\eta')G^*_{N-1}(\eta')d\eta' = H_{N-1-k}(\eta)/\Delta +$$

$$+ \sum_{p=1}^{k-1} \int_0^{2a} G_p(\eta+\eta')\, G^*_{N-1-(k-p)}(\eta')d\eta'. \quad (k \geq 2). \qquad (17)$$

Let us assume that the function $G_{N-1}(\eta')$ is known. Then Eq. (15) is linear relative to the function $G_0(\eta')$ and can serve for its determination. Equation (16) is linear relative to the functions $G_1(\eta')$ and $G_{N-2}(\eta')$. If the aperture configuration is chosen so that Eq. (16), depending on the variation interval η, is dependent only on $G_1(\eta')$ and $G_{N-2}(\eta')$, then this equation is decisive for these functions. The same is true for the left-hand sides of Eq. (17) at different k. It should be noted that the right-hand side of Eq. (17) depends on the functions $G_p(\eta')$, which are calculated from the foregoing equations.

The aperture configurations, when Eqs. (15)–(17) satisfy the above characteristics, can be constructed as follows. We subdivide the bands Ω_s into three parts $\Omega_{s,j}$, $j = 0, 1, 2$ (Fig. 1). The part $\Omega_{N-1,1}$ is denoted by ω. It is essential that some parts do not transmit the light. The remaining parts form the new aperture configuration, denoted by $W_1 + \omega$. The introduction of the region $W_0 = \Omega\setminus(W_1 + \omega)$, which does not transmit the light, is equivalent to the setting of the function $G(\xi', \eta') = 0$. Setting of a new aperture configuration is accomplished by enumerating those parts of $\Omega_{s,j}$, which appear in W_0. If the region W_0 is denoted by $W_0(N)$ at a given N, then

$$W_0(3) = \{\Omega_{N-1,0}, \Omega_{N-1,2}\},$$

$$W_0(6) = \{W_0(3), \Omega_{2,0}, \Omega_{N-2,1}\},$$

$$W_0(9) = \{W_0(6), \Omega_{3,2}\},$$

$$W_0(12) = \{W_0(9), \Omega_{5,0}, \Omega_{N-5,1}\},$$

$$W_0(15) = \{W_0(12), \Omega_{6,2}\},$$

$$W_0(18) = \{W_0(15), \Omega_{8,0}, \Omega_{N-8,1}\},$$

$$W_0(21) = (W_0(18), \Omega_{9,2}\},$$

$$W_0(24) = \{W_0(21), \Omega_{11,0}, \Omega_{N-11,1}\}, \text{ etc.}$$

The ratios of areas of the regions $W_0(N)$ and Ω are $(2 + 3n)/3N$ at $N = 3(2n + 1)$ and $(1 + 3n)/3N$ at $N = 3\cdot 2n$. At large N this ratio is close to 1/6. Figure 1 corresponds to $N = 9$.

It can be directly verified that at given aperture configurations the equations (16) and (17) are the Volterra integral equations of the first kind relative to the functions $G_k(\eta')$ and the like convolution equation relative to functions $G_{N-1-k}(\eta')$.

At a given function $G_{N-1}(\eta')$ on ω, setting sequentially $k = 0, 1, \ldots, K$, $K = (N + 1)/2$ at odd N numbers and $K = N/2$ at even N, one can find all the functions of $G_s(\eta')$ on W_1 from linear equations resulting from (16) and (17). Thus, the set of displacement vectors

$$E = \{(\xi_k, \eta) : \xi_k = -k\Delta, \quad k = 0, \ldots, K \text{ and } |\eta| \leq 2a\}$$

is sufficient for determining $G_s(\eta')$ on W_1 by the function $G_{N-1}(\eta')$ on ω. The remaining set of displacement vectors $E_0\setminus E$ can serve for determining $G_{N-1}(\eta')$ on ω from the set (17) at $k = K + 1, \ldots, N - 1$. For every $k = K + 1, \ldots, N - 1$, Eq. (17) is defined by a certain operator, transforming the functions $G_{N-1}(\eta')$. $H_{N-1-p}(\eta)(h)$, $(p = 0, 1, \ldots, K)$, and $H_{N-1-k}(\eta)$ to the function $G_{N-1}(\eta')$:

$$G_{N-1} = \varphi_k(G_{N-1}, H_{N-1}, H_{N-2}, \ldots, H_{N-1-K}, H_{N-1-k}). \qquad (18)$$

No matter how $H_{N-1-k}(\eta)$ are given, properly or with an error, we have their estimate. Therefore we can allow their small variations. Excess of the set (18) and small variations of the functions $H_{N-1-k}(\eta)$ can be used for obtaining an acceptable solution $G_{N-1}(\eta')$. The function $G_{N-1}(\eta')$ and the corresponding functions $G_s(\eta')$ on W_1 should be considered as a solution to Eq. (1).

REFERENCES

1. M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1959).
2. G. Stark, ed., *Reconstruction of Images* [Russian translation] (Mir, Moscow, 1992), 635 pp.
3. A.F. Verlan' and V.S. Sizikov, *Integral Equations: Methods, Algorithms, Programs* (Naukova Dumka, Kiev, 1986), 542 pp.
4. M. Rid and B. Saimon, *Methods of Modern Mathematical Physics*, Vol. 1, *Functional Analysis* [Russian translation] (Mir, Moscow, 1977), 357 pp.
5. B.N. Pshenichnyi, *Necessary Conditions for Extremum* (Nauka, Moscow, 1969), 151 pp.
6. S.G. Mikhlin, *Variation Methods in Mathematical Physics* (Nauka, Moscow), 512 pp.
7. P. Beits and M. Mac-Donnelli, *Restoration and Reconstruction of Images* [Russian translation] (Mir, Moscow, 1989), 333 pp.