Inversion Theory
(Lecture 6)

MICHAEL ZHDANOV

1 Formulation of well-posed and ill-posed problems

In the first lecture we have formulated an inverse problem as the solution of an operator equation

$$\mathbf{d} = A(\mathbf{m}), \tag{1}$$

where $\mathbf{m} \in M$, is some function (or vector) from a metric space M of the model parameters, and $\mathbf{d} \in D$ is an element from a metric space D of data sets. There are two important classes of inverse problems: well-posed and ill-posed problems. We will give detailed description of these problems in this section.

1.1 Well-posed problems

Following classical principles of regularization theory we can give the following definition of the well-posed problem.

Definition 1 *The problem (1) is correctly (or well) posed if the following conditions take place: (i) solution* $\mathbf{m}$ *of the equation (1) exists, (ii) solution* $\mathbf{m}$ *of the equation (1) is unique, (iii) solution* $\mathbf{m}$ *continuously depends on the right-hand side* $\mathbf{d}$.

In another words the inverse operator A^{-1} is defined throughout the space D and is continuous.

Note, that well-posed inverse problem possesses all the properties of the "good" solution discussed in the previous lectures: the solution exists, is unique and stable.

Definition 2 *The problem (1) is ill-posed if at least one of the conditions, listed above, fails.*

We have demonstrated that the majority of geophysical inverse problems are ill-posed, because at least one of the conditions listed above fails. However, it may happen that if we narrow the class of the models which are used in inversion, the originally ill-posed inverse problem may become well-posed. Mathematically it means that instead of considering $\mathbf{m}$ from entire model space M, we can select $\mathbf{m}$ from some subspace of M, consisting of more simple and/or more suitable models for the given inverse problem. Thus, we arrive at the idea of the correctness set and conditionally well-posed inverse problem.

1.2 Conditionally well-posed problems

Suppose that we know a priori that the exact solution belongs to a set C such, that the inverse operator A^{-1} defined on the image AC is continuous. We call C the *correctness set*.

Definition 3 *The problem (1) is conditionally well-posed (Tikhonov's well-posed) if the following conditions are met: (i) we know a priori that a solution of (1) exists and belongs to a specified set* $C \subset M$ *, (ii) the operator* A *is one-to-one mapping of* C *onto* $AC \subset D$*, (iii) the operator* A^{-1} *is continuous on* $AC \subset D$*.*

In contrast to the standard well-posed problems, a conditionally well-posed problem doesn't require solvability over the whole space. Also the requirement of continuity of A^{-1} over all M is substituted by the requirement of continuity over the image of C. Thus, introducing a correctness set makes even ill-posed problem to be well-posed, if we reduce the class of possible solutions.

Tikhonov described the correctness set C. For example , if the model is described by a finite number of bounded parameters, they form a correctness set C in Euclidean space of model parameters.

1.3 Quasi-solution of ill-posed problem

We assume now that the problem (1) is conditionally well-posed (Tikhonov's well-posed). Suppose also that the right-hand side of (1) is given with some errors:

$$\mathbf{d}_\delta = \mathbf{d} + \delta\mathbf{d}, \tag{2}$$

where

$$\mu_D(\mathbf{d}_\delta, \mathbf{d}) \leq \delta \tag{3}$$

Definition 4 *A quasi-solution of the problem (1) on the correctness set C is an element $\mathbf{m}_\delta \in C$ which minimizes the distance $\mu_D(A\mathbf{m}, \mathbf{d}_\delta)$, i.e.:*

$$\mu_D(A\mathbf{m}_\delta, \mathbf{d}_\delta) = \inf_{m \in C \mu D} \mu_D(A\mathbf{m}, \mathbf{d}_\delta). \tag{4}$$

It can be proved also, that the quasi-solution is a continuous function of $\mathbf{d}_\delta$.

Figure 1 illustrates the definition of a quasi-solution. Element $\mathbf{m} \in \mathbf{M}$ is exact solution of the inverse problem

$$\mathbf{d} = A(\mathbf{m}). \tag{5}$$

Subset AC of the data space D is an image of the correctness set C as the result of operator A applications. A quasi-solution, $\mathbf{m}_\delta$, is selected from a correctness set C under the condition that its image, $A(\mathbf{m}_\delta)$, is the closest element to the observed noisy data, $\mathbf{d}_\delta$,from subset AC.

The idea of quasi-solution makes it possible to substitute the inverse problem solution by minimization of the distance $\mu_D(A\mathbf{m}, \mathbf{d}_\delta)$ on some appropriate class of suitable models. The standard methods of functional minimization can be used to solve this problem and therefore to find the quasi-solution. In this way, we significantly simplify the inverse problem solution.

2 The regularization method in the solution of the inverse problem

2.1 Regularizing operators

Consider inverse geophysical problem described by the operator equation:

$$\mathbf{d} = A(\mathbf{m}), \tag{6}$$

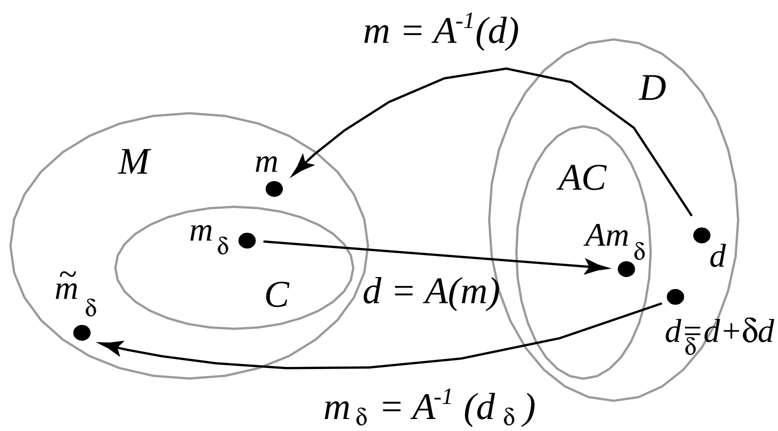

Figure 1: A quasi-solution, $\mathbf{m}_\delta$, is selected from the correctness set C under the condition that its image, $A(\mathbf{m}_\delta)$, is the closest element to the observed noisy data, $\mathbf{d}_\delta$, from the subset AC : $\mu_D(A\mathbf{m}_\delta, \mathbf{d}_\delta) = \inf_{m \in C} \mu_D(A\mathbf{m}, \mathbf{d}_\delta)$

where $\mathbf{m}$ presents model parameters and $\mathbf{d}$ is observed geophysical data. In general case the inverse operator A^{-1} is not continuous and therefore the inverse problem (6) is ill-posed. The main idea of any regularization algorithm is to consider instead of one ill-posed inverse problem (6) a family of the well-posed problems

$$\mathbf{d} = A_\alpha(\mathbf{m}), \tag{7}$$

which approximate in some sense the original inverse problem. Scalar parameter $\alpha > 0$ is called *a regularization parameter*. We require also that

$$\mathbf{m}_\alpha \rightarrow \mathbf{m}_t, \ if \ \alpha \rightarrow 0,$$

where $\mathbf{m}_\alpha = A_\alpha^{-1}(\mathbf{d})$ is the solution of the inverse problems (7), and $\mathbf{m}_t$ is the true solution of the original problem (6). Thus, we substitute the solution of one ill-posed inverse problem by the solutions of the family of well-posed problems, assuming that these solutions, $\mathbf{m}_\alpha$, asymptotically go to the true solution if α goes to zero.

Let us give now the more accurate definitions.

Definition 5 *Operator $R(\mathbf{d}, \alpha)$ (depending on a scalar parameter α) is called the regularizing operator in some vicinity of the element $\mathbf{d}_t = A(\mathbf{m}_t)$ if there is a function $\alpha(\delta)$ such that for any $\epsilon > 0$ it can be found a positive number $\delta(\epsilon)$ such that if*

$$\mu_D(\mathbf{d}, \mathbf{d}_t) < \delta(\epsilon),$$

then

$$\mu_M(\mathbf{m}_\alpha, \mathbf{m}_t) < \epsilon,$$

where

$$\mathbf{m}_\alpha = R(\mathbf{d}, \alpha(\delta)).$$

In other words $\mathbf{m}_\alpha$ is a continuous function of the data and

$$\mathbf{m}_\alpha = R(\mathbf{d}, \alpha(\delta)) \rightarrow \mathbf{m}_t \tag{8}$$

when $\alpha \rightarrow 0$.

Figure 2 illustrates the basic properties of the regularizing operators. Let $\mathbf{m}_t$ is exact solution for exact data $\mathbf{d}_t = A(\mathbf{m}_t)$. However, we can observe only the noisy data $\mathbf{d}_\delta = \mathbf{d}_t + \delta\mathbf{d}$. If we apply some rigorous inverse operator to noisy data, we could get a result $\mathbf{m}'_\delta$ which lies far away from true solution.

By perturbing noisy data $\widetilde{\mathbf{d}}_\delta$ one can get another completely different from the true solution result $\widetilde{\mathbf{m}}'_\delta$. The main advantage of the regularizing operators R is that they provide the stable solution in any situation. If we apply R to the noisy data $\mathbf{d}_\delta$, we will get a solution $\mathbf{m}_\delta = R(\mathbf{d}_\delta, \alpha)$ which is very close to a true model $\|\mathbf{m}_\delta - \mathbf{m}_t\| < \epsilon$. Application of R to a noisy data $\widetilde{\mathbf{d}}_\delta$ will result in another solution $\widetilde{\mathbf{m}}_\delta = R(\widetilde{d}_\delta, \alpha)$, which is still close to $\mathbf{m}_t$. The accuracy of the true solution approximation by the regularized one depends on the regularization parameter α. The smaller the α the more accurate is the approximation.

We can see that regularizing operators can be constructed by approximating ill-posed equation (6) by the system of well-posed equations (7), where corresponding inverse operators A_α^{-1} are continuous. These inverse operators can be treated as regularizing operators

$$A_\alpha^{-1}(\mathbf{d}) = R(\mathbf{d}, \alpha).$$

The only problem now is how to find the family of the regularizing operators. Tikhonov suggested the following scheme for constructing regularizing operators. It is based on introducing special stabilizing and parametric functionals.

2.2 Stabilizing functional

Stabilizing functional (or stabilizer) is used to select from the space M of all possible models the subset M_c which is a correctness set.

Definition 6 *A nonnegative functional $s(\mathbf{m})$ on some metric space M is called a stabilizing functional if for any real number $c > 0$ the subset M_c of elements $\mathbf{m} \in M$ for which $s(\mathbf{m}) \leq c$ is a correctness set.*

We will give now an example of the stabilizing functionals.

Example 7 *Consider the Hilbert space L_2 formed by the functions integrable on the interval $[a, b]$. The metric in the space L_2 is determined according to formula:*

$$\mu(\mathbf{m}_1, \mathbf{m}_2) = \left\{ \int_a^b [m_1(x) - m_2(x)]^2 \, dx \right\}^{\frac{1}{2}}, \tag{9}$$

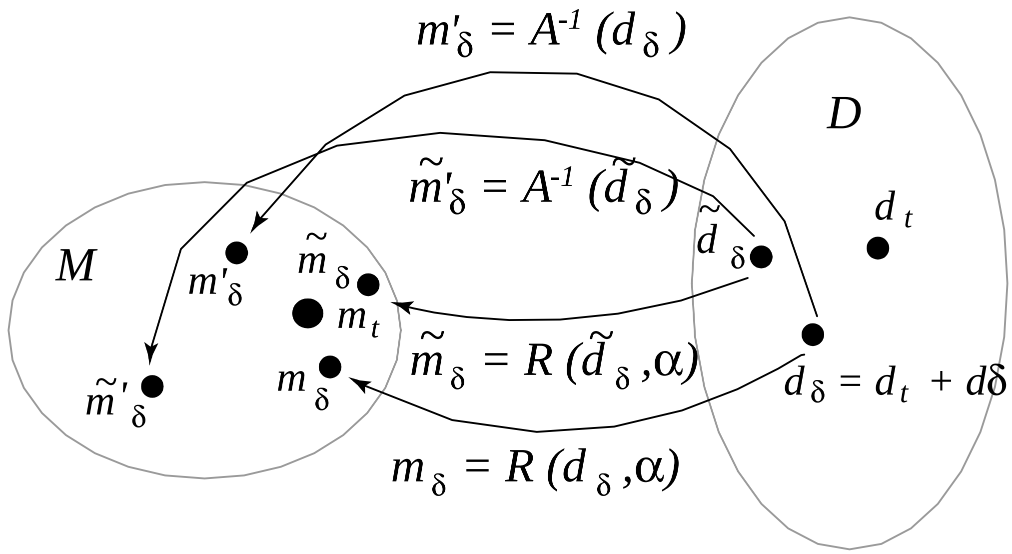

Figure 2: The scheme illustrating the construction of the regularizing operators. The bold point $\mathbf{m}_t$ denotes the true solution, and the point $\mathbf{d}_t \in D$ denotes the true data. The noisy data are shown by a point $\mathbf{d}_\delta = \mathbf{d}_t + \delta\mathbf{d}$. Application of a formal inverse operator A^{-1} to the noisy data generates formal solutions, $\mathbf{m}'_\delta$ and $\widetilde{\mathbf{m}}'_\delta$, which are unstable with respect to a small perturbation in the data, $\mathbf{d}_\delta$ or $\widetilde{\mathbf{d}}_\delta$. However, application of the regularizing operators, $R(\mathbf{d}, \alpha)$ to any of the observed data, $\mathbf{d}_\delta$ or $\widetilde{\mathbf{d}}_\delta$, produces stable results: the inverse models $\mathbf{m}_\delta$ and $\widetilde{\mathbf{m}}_\delta$ are close to each other and to the true solution, if the observed data, $\mathbf{d}_\delta$ and $\widetilde{\mathbf{d}}_\delta$, are close to each other and to the true data $\mathbf{d}_t$.

It can be proved that any ball

$$b\,(\mathbf{m}_0,c) = \{\mathbf{m} : \mu(\mathbf{m},\mathbf{m}_0) \le c\,\}$$

is a correctness set in a Hilbert space. Therefore, we can introduce the stabilizing functional as following:

$$s(\mathbf{m}) = \mu(\mathbf{m},\mathbf{m}_0), \tag{10}$$

where $\mathbf{m}_0$ *is any given model from* $M = L_2$ *and a ball :*

$$s(m) = \mu(m, m_0) \le c, \tag{11}$$

is a correctness set.

Let us analyze now more carefully how one can use the stabilizers to select an appropriate class of the models. Assume that the data $\mathbf{d}_\delta$ are observed with some noise $\mathbf{d}_\delta = \mathbf{d}_t + \delta\mathbf{d}$, where $\mathbf{d}_t$ is the true solution of the problem. In other words, we assume that the misfit (distance) between the observed data and true data is less than the given level of the errors in the observed data, equal to $\delta \ge \|\delta\mathbf{d}\|$:

$$\mu_D(\mathbf{d}_\delta, \mathbf{d}_t) \le \delta. \tag{12}$$

In this situation it is naturally to search for an approximate solution in the set Q_δ of the models $\mathbf{m}$ such that

$$\mu_D(A(\mathbf{m}), \mathbf{d}_\delta) \le \delta. \tag{13}$$

Thus $Q_\delta \subset M$ is the set of possible solutions.

Figure 3 helps to understand this role of the stabilizing functional. The main application of stabilizers is to select from the set of possible solutions Q_δ the solutions, which continuously depend on the data and which possesses a specific property depending on the choice of a stabilizer. Such solutions can be selected by the condition of the minimum of the stabilizing functional:

$$s(\mathbf{m};\ \mathbf{m} \in Q_\delta) = min. \tag{14}$$

We have introduced a stabilizing functional under the condition that it selects a correctness subset M_C from a metric space of the model parameters. Thus,

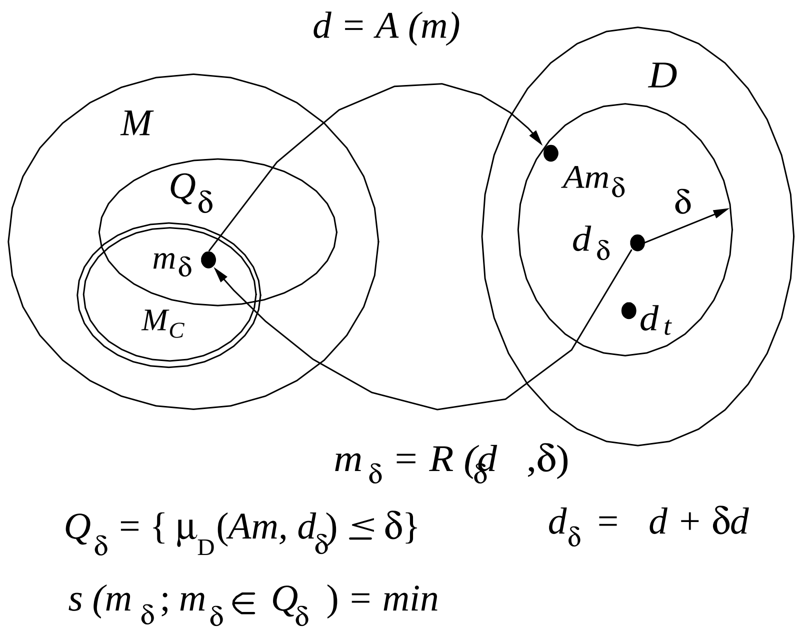

Figure 3: The stabilizing functional selects from a set of the possible solutions, Q_δ, a solution, $\mathbf{m}_\delta$, which at the same time belongs to the correctness set M_C.

we can say that a stabilizer selects from the set of possible solutions Q_δ a solution, which at the same time belongs to a correctness set M_C.

The existence of the model, minimizing (14), was demonstrated in (Tikhonov and Arsenin, 1977). We will denote this model as $\mathbf{m}_\delta$:

$$s(\mathbf{m}_\delta;\ \mathbf{m}_\delta \in Q_\delta) = min. \tag{15}$$

One can consider a model $\mathbf{m}_\delta$ as the result of application to the observed data $\mathbf{d}_\delta$ an operator $R(\mathbf{d}_\delta, \delta)$, depending on the parameter δ:

$$\mathbf{m}_\delta = R(\mathbf{d}_\delta, \delta) \tag{16}$$

It can be proved that the operator $R(\mathbf{d}_\delta, \delta)$ is the regularizing operator for the equation (**??**) and $\mathbf{m}_\delta$ can be used as the approximate solution of the inverse problem (note that in this case $\alpha = \delta$, while in general case $\alpha = \alpha(\delta)$).

Thus in the framework of developed approach the problem of the solution of eq. (1) with approximate left hand part $\mathbf{d}_\delta$ can be reduced to the problem of minimization of stabilizing functional on the set Q_δ:

$$s(\mathbf{m};\ \mathbf{m} \in Q_\delta) = min,$$

where

$$Q_\delta = \{\mathbf{m}; \mu_D(A(\mathbf{m}), \mathbf{d}_\delta) \leq \delta\}. \tag{17}$$

2.3 Tikhonov parametric functional

It can be proved that for a wide class of stabilizing functionals they reach their minimum on the model $\mathbf{m}_\delta$ such that $\mu_D(A(\mathbf{m}_\delta), \mathbf{d}_\delta) = \delta$. Thus , we can solve the problem of minimization (14) under the condition that

$$\mu_D(A(\mathbf{m}_\delta), \mathbf{d}_\delta) = \delta. \tag{18}$$

The last problem is the problem of the stabilizing functional (14) minimization when the model $\mathbf{m}$ is subject to the constrain (18). A common way to solve this problem is to introduce an unconstrained functional $P^\alpha(\mathbf{m}, \mathbf{d}_\delta),\ \mathbf{m} \in M$, given by

$$P^\alpha(\mathbf{m}, \mathbf{d}_\delta) = \mu_D^2(A(\mathbf{m}), d_\delta) + \alpha s(\mathbf{m}), \tag{19}$$

and to solve the problem of minimization of this functional:

$$P^{\alpha}(\mathbf{m}, \mathbf{d}_{\delta}) = \min. \tag{20}$$

Here the unknown real parameter α is similar to the Lagrangian multiplier and it is determined under the condition

$$\rho_D(A(\mathbf{m}_{\alpha}), d_{\delta}) = \delta, \tag{21}$$

where $\mathbf{m}_{\alpha}$ is the element on which $P^{\alpha}(\mathbf{m}, \mathbf{d}_{\delta})$ reaches minimum. The functional $P^{\alpha}(\mathbf{m}, \mathbf{d}_{\delta})$ is called *Tikhonov parametric functional.*

Thus for any positive number $\alpha > 0$ and for any data $\mathbf{d}_{\delta} \in D$ we have determined an operator $R(\mathbf{d}_{\delta}, \alpha)$ with the values in M such that the model

$$\mathbf{m}_{\alpha} = R(\mathbf{d}_{\delta}, \alpha) \tag{22}$$

gives the minimum of the Tikhonov parametric functional $P^{\alpha}(\mathbf{m}, \mathbf{d}_{\delta})$.

The fundamental result of the regularization theory is that this *operator $R(\mathbf{d}_{\delta}, \alpha)$ is the regularizing operator for the problem (1).*Thus as an approximate solution of the inverse problem (1) we take the solution of another problem (20) (problem of minimization of Tikhonov parametric functional $P^{\alpha}(\mathbf{m}, \mathbf{d}_{\delta})$) , closed to initial problem for the small values of the errors δ of the data $\mathbf{d}_{\delta}$.

It is important to underline that in the case when A is a linear operator, D and M are Hilbert spaces and $s(\mathbf{m})$ is a quadratic functional , the solution of the minimization problem (**??**) is unique. Note that the quadratic functional is a functional $q(\mathbf{m})$ such, that

$$q(\beta \mathbf{m}) = \beta^2 q(\mathbf{m}).$$

2.4 Determination of the regularization parameter

One of the most critical problems of regularization method is the selection of the optimum value of the regularization parameter α. The solution of this problem can be based on the following consideration.

Suppose that the data $\mathbf{d}_{\delta}$ is observed with some noise $\mathbf{d}_{\delta} = \mathbf{d}_t + \delta \mathbf{d}$, where $\mathbf{d}_t$ is the true solution of the problem, and the level of the errors in the observed data is equal to δ:

$$\mu_D(\mathbf{d}_{\delta}, \mathbf{d}_t) \leq \delta. \tag{23}$$

Then the regularization parameter can be determined by the misfit condition (22)

$$\mu_D(A(\mathbf{m}_\alpha), \mathbf{d}_\delta) = \delta. \tag{24}$$

To justify this approach we will examine more carefully the properties of all three functionals involved in the regularization method: the Tikhonov parametric functional, the stabilizing and misfit functionals.

Let us introduce the following notations:

$$p(\alpha) = P^\alpha(\mathbf{m}_\alpha, \mathbf{d}_\delta);$$

$$i(\alpha) = \mu_D(A(\mathbf{m}_\alpha), \mathbf{d}_\delta); \tag{25}$$

$$s(\alpha) = s(\mathbf{m}_\alpha).$$

Examine some properties of the functions $p(\alpha), i(\alpha), s(\alpha)$.

Property 1

Functions $p(\alpha), i(\alpha), s(\alpha)$ are monotone functions: $p(\alpha)$ and $i(\alpha)$ are not decreasing and $s(\alpha)$ is not increasing.

Property 2

It can be proved that the functions $p(\alpha), i(\alpha), s(\alpha)$ are continuous functions (if the element m_α is unique).

Note also that

$$p(\alpha) \rightarrow 0 \ for \ \alpha \rightarrow 0,$$

and

$$p(0) = 0. \tag{26}$$

From the fact that

$$i(\alpha) + \alpha s(\alpha) = p(\alpha) \rightarrow 0 \ for \ \alpha \rightarrow 0,$$

it follows that

$$i(0) = 0. \tag{27}$$

Thus one can prove the following theorem.

Theorem 8 *If $i(\alpha)$ is one-to-one function then for any positive number $\delta < \delta_0 = \mu_D(A(\mathbf{m}_0), \mathbf{d}_\delta)$ (where $\mathbf{m}_0$ is some a priori model) there exists $\alpha(\delta)$ such that $\mu_D(A(\mathbf{m}_{\alpha(\delta)}), \mathbf{d}_\delta) = \delta$*

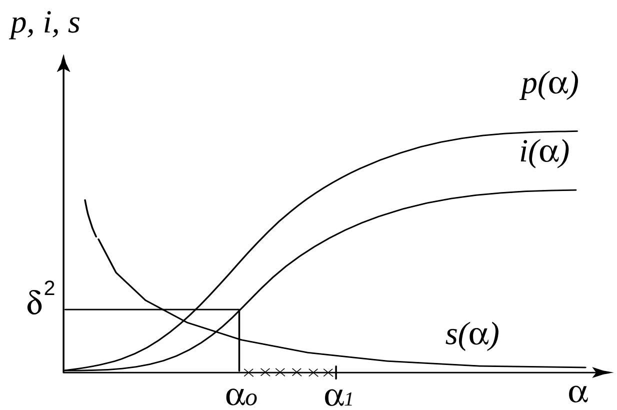

Figure 4: Illustration of the principle of optimal regularization parameter selection

Note that $i(\alpha)$ is one-to-one function when element $\mathbf{m}_\alpha$ is unique. It happens, for example, when A is a linear operator, D is a Hilbert space and $s(\mathbf{m})$ is a quadratic functional.

Figure 4 helps to understand the principle of optimal regularization parameters selection. One can see that due to monotone character of the function $s(\alpha)$, there is only one point, α_0, where $s(\alpha_0) = \delta$. Let us consider one simple numerical method of determination of parameter α. Consider for example the progression of numbers:

$$\alpha_k = \alpha_0 q^k, \; k = 1, 2, ..., n; \; q > 0. \tag{28}$$

For any number α_k we can find element $\mathbf{m}_{\alpha_k}$, minimizing $P^{\alpha_k}(\mathbf{m}, \mathbf{d}_\delta)$ and calculate the misfit $\mu_D(A(\mathbf{m}_{\alpha_k}), \mathbf{d}_\delta)$. The optimal value of the parameter α is the number α_{k0}, for which with the necessary accuracy we have the equality

$$\mu_D(A(\mathbf{m}_{\alpha_{k0}}), \mathbf{d}_\delta) = \delta \tag{29}$$

The equality (29) is called the *misfit condition.*

2.5 L-curve method of regularization parameter selection

L-curve analysis (Hansen, 1998) represents a simple graphical tool for qualitative selection of the quasi-optimal regularization parameter. It is based on plotting for all possible α the curve of the misfit functional, $i(\alpha)$,versus the stabilizing functional, $s(\alpha)$ (where we use notations (25)). The L-curve illustrates the trade-off between the best fitting (minimizing a misfit) and most reasonable stabilization (minimizing a stabilizer). In a case where α is selected to be too small, the minimization of the parametric functional $P^{\alpha}(\mathbf{m})$ is equivalent to the minimization of the misfit functional; therefore $i(\alpha)$ decreases, while $s(\alpha)$ increases. When α is too large, the minimization of the parametric functional $P^{\alpha}(\mathbf{m})$ is equivalent to the minimization of the stabilizing functional; therefore $s(\alpha)$ decreases, while $i(\alpha)$ increases. As a result, it turns out that the L-curve, when it is plotted in log-log scale, very often has a characteristic L-shape appearance (Figure 5), that justifies its name (Hansen, 1998).

The distinct corner, separating the vertical and the horizontal branches of this curve, corresponds to the quasi-optimal value of the regularization parameter α.

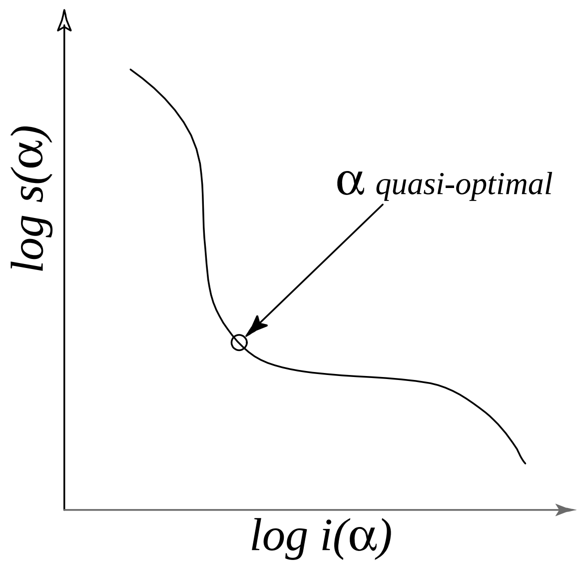

Figure 5: L-curve represents a simple curve for all possible α of the misfit functional, $i(\alpha)$, versus stabilizing functional, $s(\alpha)$, plotted in log-log scale. The distinct corner, separating the vertical and the horizontal branches of this curve, corresponds to the quasi-optimal value of the regularization parameter α.