

Generalized Adaptive versus Nonadaptive Recovery from Noisy Information

K. Yu. Osipenko¹

*Moscow State University,
Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow*

Abstract

The article is devoted to a generalized setting of adaptive recovery from noisy information. We extend some known results for this setting and obtain sufficient conditions under which adaption does not help for recovery of linear operators. For recovery of linear functionals we give necessary and sufficient conditions for the existence of an optimal nonadaptive linear method which gives the same error as optimal adaptive methods.

Keywords: optimal recovery, adaptive and nonadaptive methods, adaptive radius of information

2010 MSC: 41A65, 41A45, 41A46, 49N30

1. Introduction

Let X be a linear space, Z be a normed linear space, $T: X \rightarrow Z$ be a linear operator and $W \subset X$. We consider the problem of recovery of T on the set $W \subset X$ by noisy information about elements from W . Suppose that there is a family \mathcal{I} of linear information operators $I_p: X \rightarrow Y$, $p \in \Omega$, where Ω is some set and Y is a normed linear space. All linear spaces are considered over the field of real or complex numbers. Let us choose n parameters $p_1, \dots, p_n \in \Omega$ and assume that for all $x \in W$ we know $y_1, \dots, y_n \in Y$ such that

$$\|I_{p_j}x - y_j\|_Y \leq \delta, \quad \delta \geq 0.$$

As recovery methods we consider all possible mappings $\varphi: Y^n \rightarrow Z$. The error of a method φ is defined as follows

$$e_n(T, W, \mathcal{I}, \delta, \mathbf{p}, \varphi) = \sup_{\substack{x \in W, \mathbf{y}=(y_1, \dots, y_n) \in Y^n \\ \|I_{p_j}x - y_j\|_Y \leq \delta, j=1, \dots, n}} \|Tx - \varphi(\mathbf{y})\|_Z;$$

Email address: kosipenko@yahoo.com (K. Yu. Osipenko)

¹The research was carried out with the financial support of the Russian Foundation for Basic Research (grant no. 17-01-00649)

here $\mathbf{p} = (p_1, \dots, p_n)$. The quantity

$$E_n(T, W, \mathcal{I}, \delta) = \inf_{p_1, \dots, p_n \in \Omega} \inf_{\varphi: Y^n \rightarrow Z} e_n(T, W, \mathcal{I}, \delta, \mathbf{p}, \varphi)$$

is called the n th optimal nonadaptive recovery error. Parameters $\hat{p}_1, \dots, \hat{p}_n$ and the method $\hat{\varphi}$ we call optimal if

$$e_n(T, W, \mathcal{I}, \delta, \hat{\mathbf{p}}, \hat{\varphi}) = E_n(T, W, \mathcal{I}, \delta), \quad \hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_n).$$

One of the typical examples of this setting is the problem of optimal integration. Let W be a class of functions defined on the segment $[a, b]$. Set

$$Tx = \int_a^b x(t) dt, \quad I_{t_j}x = x(t_j), \quad t_j \in [a, b], \quad j = 1, \dots, n.$$

Then the problem stated above becomes the problem of choosing optimal method of integration and optimal nodes of integration at which the function should be measured.

In practice, quite often a different approach is used, in which subsequent nodes of integration are defined on the base of information about function values at the previous nodes. Such algorithms are called adaptive (or sequential). In this regard, it makes sense to consider a slightly different (more general) setting of the optimal recovery problem.

Let $\omega_1 \in \Omega$ and functions

$$\omega_j: Y^{j-1} \rightarrow \Omega, \quad j = 2, \dots, n,$$

be given. In case of adaptive information, we assume that for $x \in W$ we know $\mathbf{y} = (y_1, \dots, y_n)$ such that

$$\|I_{p_j}x - y_j\|_Y \leq \delta, \quad j = 1, \dots, n,$$

where $p_1 = \omega_1$ and

$$p_j = \omega_j(y_1, \dots, y_{j-1}), \quad j = 2, \dots, n. \quad (1)$$

As recovery methods we consider all possible mappings

$$\varphi: Y^n \rightarrow Z.$$

For $x \in W$ we set

$$\Lambda(x) = \{ (y_1, \dots, y_n) : \|I_{p_j}x - y_j\|_Y \leq \delta, \quad j = 1, \dots, n \},$$

where $p_1 = \omega_1$ and $p_j, j = 2, \dots, n$, are defined by (1), that is, $\Lambda(x)$ is the set of all possible information \mathbf{y} about x .

The error of a method φ is defined as follows

$$e_n^a(T, W, \mathcal{I}, \delta, \omega, \varphi) = \sup_{\substack{x \in W \\ \mathbf{y} = (y_1, \dots, y_n) \in \Lambda(x)}} \|Tx - \varphi(\mathbf{y})\|_Z;$$

here $\omega = (\omega_1, \dots, \omega_n)$. The value

$$E_n^a(T, W, \mathcal{I}, \delta) = \inf_{\omega, \varphi} e_n(T, W, \mathcal{I}, \delta, \omega, \varphi)$$

is called the error of optimal adaptive recovery.

Adaptive methods have a much richer structure and it is natural to expect that the error of adaptive recovery will be smaller than for nonadaptive recovery. This is what happens in a number of problems, such as calculating the root of a function or finding the maximum for an unimodal function (note that in these examples, the operator to be recovered is not linear). However, it turns out that there are cases when adaptive recovery does not give a win. In such cases, instead of complicated adaptive methods, it makes sense to use simpler nonadaptive methods.

The question of whether adaption helps to reduce the error of recovery has been studied in various papers. In [1] it was proved that adaptive methods do not help when T is a linear functional and W is a convex and centrally-symmetric set. It was shown in [2] and [15] (see also [13]) that adaption helps at most twice for recovery of linear operators T over convex and symmetric sets W (we extend this result for the generalized adaption in Theorem 2). In [1], [2], and [15] only noiseless situation was considered. An example of such problem (with exact information) where adaption can help was given in [3], [4]. It was shown in [5], [9], [10] that for convex but not symmetric sets adaptive information can be significantly better than nonadaptive information.

The results on adaption versus nonadaption in the average case setting may be found in [16] and [12] (see also [14] and [13]).

In the present paper we give a generalized setting of adaptive recovery from noisy information and extend some known results for this setting. Our main results are sufficient conditions under which adaption does not help for recovery of linear operators (Theorem 3), an application of this result to recovery of derivatives by noisy information about the Fourier coefficients (Theorem 4), and criterion for the existence of an optimal nonadaptive linear method which gives the same error as optimal adaptive methods for recovery of linear functionals (Theorem 5).

2. Generalized setting of adaptive recovery

In adaptive recovery, we can define mappings

$$\Phi(\mathbf{y}) = (\omega_1, \omega_2(y_1), \dots, \omega_n(y_1, \dots, y_{n-1})) \quad (2)$$

which for every $\mathbf{y} = (y_1, \dots, y_n)$ give the values of parameters $\mathbf{p} = (p_1, \dots, p_n)$ by the formula $\mathbf{p} = \Phi(\mathbf{y})$. We now generalize the adaptive setting by allowing arbitrary mappings $\Phi(\mathbf{y})$.

We will consider the following problem of optimal adaptive recovery. Let X , Z , T , and W be as above and \mathcal{Y} be a linear space. Assume that there is a set of parameters \mathcal{P} and for every $\mathbf{p} \in \mathcal{P}$ a multivalued mapping

$$F_{\mathbf{p}}: W \rightarrow \mathcal{Y}$$

is assigned (that is, for every $x \in W$ $F_{\mathbf{p}}(x)$ is a set from \mathcal{Y}). The mapping $F_{\mathbf{p}}$ represents the nonadaptive information corresponding to the indices $\mathbf{p} \in \mathcal{P}$. Denote by \mathcal{F} the set of all mappings $F_{\mathbf{p}}$.

Suppose also that there is a mapping

$$\Phi: \mathcal{Y} \rightarrow \mathcal{P}.$$

As recovery methods of T we consider all possible mappings $\varphi: \mathcal{Y} \rightarrow Z$.

The error of a method φ is defined as follows

$$e^a(T, \mathcal{F}, \Phi, \varphi) = \sup_{\substack{x \in W \\ \mathbf{y} \in F_{\Phi(\mathbf{y})}(x)}} \|Tx - \varphi(\mathbf{y})\|_Z,$$

and the value

$$E^a(T, \mathcal{F}) = \inf_{\Phi: \mathcal{Y} \rightarrow \mathcal{P}} \inf_{\varphi: \mathcal{Y} \rightarrow Z} e^a(T, \mathcal{F}, \Phi, \varphi)$$

is called the error of optimal adaptive recovery.

In this setting we do not see the sequential procedure in the explicit form. Nevertheless sequential procedures are included in the set of all mappings Φ (it suffices to take Φ in the form (2)).

By the nonadaptive problem of recovery for the considered case we mean the problem of finding the value

$$E(T, \mathcal{F}) = \inf_{\mathbf{p} \in \mathcal{P}} \inf_{\varphi: \mathcal{Y} \rightarrow Z} e(T, \mathcal{F}, \mathbf{p}, \varphi), \quad (3)$$

where

$$e(T, \mathcal{F}, \mathbf{p}, \varphi) = \sup_{\substack{x \in W \\ \mathbf{y} \in F_{\mathbf{p}}(x)}} \|Tx - \varphi(\mathbf{y})\|_Z,$$

and also the problem of finding parameters and methods for which the lower bound is attained.

If $\Phi(\mathbf{y}) = \mathbf{p}$ for all $\mathbf{y} \in \mathcal{Y}$, then

$$e^a(T, \mathcal{F}, \Phi, \varphi) = e(T, \mathcal{F}, \mathbf{p}, \varphi).$$

Therefore,

$$E^a(T, \mathcal{F}) \leq E(T, \mathcal{F}). \quad (4)$$

3. Some relations between adaptive and nonadaptive recovery errors

Set

$$\text{gr } F_{\mathbf{p}} := \{ (x, \mathbf{y}) : x \in W, \mathbf{y} \in F_{\mathbf{p}}(x) \}$$

and

$$F_{\mathbf{p}}^{-1}(\mathbf{y}) = \{ x \in W : (x, \mathbf{y}) \in \text{gr } F_{\mathbf{p}} \}.$$

Put

$$D(\mathbf{p}) = \sup_{x \in F_{\mathbf{p}}^{-1}(0)} \|Tx\|_Z.$$

Theorem 1. Assume that for all $\mathbf{p} \in \mathcal{P}$ the sets $\text{gr } F_{\mathbf{p}}$ are centrally-symmetric and contain 0. Then

$$E^a(T, \mathcal{F}) \geq \inf_{\mathbf{p} \in \mathcal{P}} D(\mathbf{p}).$$

Proof. It follows from the definition of $E^a(T, \mathcal{F})$ that for any $\varepsilon > 0$ there exist $\widehat{\Phi}$ and $\widehat{\varphi}$ such that

$$e^a(T, \mathcal{F}, \widehat{\Phi}, \widehat{\varphi}) \leq E^a(T, \mathcal{F}) + \varepsilon.$$

Consider the value $D(\widehat{\Phi}(0))$. For any $\eta > 0$ there exists a $\widehat{x} \in F_{\widehat{\Phi}(0)}^{-1}(0)$ such that

$$\|T\widehat{x}\|_Z \geq D(\widehat{\Phi}(0)) - \eta.$$

Since the set $\text{gr } F_{\widehat{\Phi}(0)}$ is centrally-symmetric $-\widehat{x} \in F_{\widehat{\Phi}(0)}^{-1}(0)$. We have

$$\begin{aligned} 2(D(\widehat{\Phi}(0)) - \eta) &\leq 2\|T\widehat{x}\|_Z = \|T\widehat{x} - \widehat{\varphi}(0) - (T(-\widehat{x}) - \widehat{\varphi}(0))\|_Z \\ &\leq \|T\widehat{x} - \widehat{\varphi}(0)\|_Z + \|T(-\widehat{x}) - \widehat{\varphi}(0)\|_Z \leq 2e^a(T, \mathcal{F}, \widehat{\Phi}, \widehat{\varphi}) \\ &\leq 2(E^a(T, \mathcal{F}) + \varepsilon). \end{aligned}$$

This implies that

$$D(\widehat{\Phi}(0)) \leq E^a(T, \mathcal{F}).$$

Consequently,

$$\inf_{\mathbf{p} \in \mathcal{P}} D(\mathbf{p}) \leq E^a(T, \mathcal{F}).$$

□

Put

$$e(T, \mathcal{F}, \mathbf{p}) = \inf_{\varphi: \mathcal{Y} \rightarrow Z} e(T, \mathcal{F}, \mathbf{p}, \varphi).$$

Lemma 1. Assume that for all $\mathbf{p} \in \mathcal{P}$ the sets $\text{gr } F_{\mathbf{p}}$ are convex and centrally-symmetric. Then for all $\mathbf{p} \in \mathcal{P}$

$$e(T, \mathcal{F}, \mathbf{p}) \leq 2D(\mathbf{p}).$$

Proof. Denote by $\text{Pr}_{\mathcal{Y}} \text{gr } F_{\mathbf{p}}$ the set of $\mathbf{y} \in \mathcal{Y}$ for which $F_{\mathbf{p}}^{-1}(\mathbf{y}) \neq \emptyset$. If $\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{gr } F_{\mathbf{p}}$ and $x_1, x_2 \in F_{\mathbf{p}}^{-1}(\mathbf{y})$ then in view of convexity and centrally-symmetry of $\text{gr } F_{\mathbf{p}}$ $h = (x_1 - x_2)/2 \in F_{\mathbf{p}}^{-1}(0)$. Set

$$\varphi_0(\mathbf{y}) = \begin{cases} Tx(\mathbf{y}), & \mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{gr } F_{\mathbf{p}}, \\ 0, & \mathbf{y} \notin \text{Pr}_{\mathcal{Y}} \text{gr } F_{\mathbf{p}}, \end{cases}$$

where $x(\mathbf{y})$ is any element from $F_{\mathbf{p}}^{-1}(\mathbf{y})$. Then for all $\mathbf{y} \in \mathcal{Y}$

$$\sup_{x \in F_{\mathbf{p}}^{-1}(\mathbf{y})} \|Tx - \varphi_0(\mathbf{y})\|_Z \leq 2\|Th\|_Z \leq 2D(\mathbf{p}),$$

where $h = (x - x(\mathbf{y}))/2 \in F_{\mathbf{p}}^{-1}(0)$. Thus,

$$e(T, \mathcal{F}, \mathbf{p}, \varphi_0) = \sup_{\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } F_{\mathbf{p}}} \sup_{x \in F_{\mathbf{p}}^{-1}(\mathbf{y})} \|Tx - \varphi_0(\mathbf{y})\|_Z \leq 2D(\mathbf{p}).$$

Consequently, for all $\mathbf{p} \in \mathcal{P}$

$$e(T, \mathcal{F}, \mathbf{p}) \leq e(T, \mathcal{F}, \mathbf{p}, \varphi_0) \leq 2D(\mathbf{p}).$$

□

In view of (4) from Theorem 1 and Lemma 1 we have

Theorem 2. *If the sets $\text{gr } F_{\mathbf{p}}$ are convex and centrally-symmetric for all $\mathbf{p} \in \mathcal{P}$, then*

$$\frac{1}{2}E(T, \mathcal{F}) \leq E^a(T, \mathcal{F}) \leq E(T, \mathcal{F}).$$

Theorem 3. *Assume that for all $\mathbf{p} \in \mathcal{P}$ the sets $\text{gr } F_{\mathbf{p}}$ are centrally-symmetric, contain 0, and the equality*

$$e(T, \mathcal{F}, \mathbf{p}) = D(\mathbf{p}) \tag{5}$$

holds. Then

$$E^a(T, \mathcal{F}) = E(T, \mathcal{F}). \tag{6}$$

Proof. It follows from Theorem 1 that

$$E(T, \mathcal{F}) = \inf_{\mathbf{p} \in \mathcal{P}} D(\mathbf{p}) \leq E^a(T, \mathcal{F}).$$

Taking into account (4) we obtain (6). □

Thus, in the case when condition (5) holds optimal nonadaptive methods give the same error as optimal adaptive methods. In other words, in this case adaption does not help.

Condition (5) is valid when T is a linear functional (see [6]). But in some cases this condition may be also valid for linear operators. We give one of such examples.

Let \mathbb{T} be the interval $[-\pi, \pi]$ with identified endpoints. Denote by $\mathcal{W}_2^r(\mathbb{T})$ the Sobolev space of 2π -periodic functions with absolutely continuous the $(r-1)$ st derivative such that the r th derivative is in $L_2(\mathbb{T})$. Set

$$W_2^r(\mathbb{T}) = \{x(\cdot) \in \mathcal{W}_2^r(\mathbb{T}) : \|x^{(r)}(\cdot)\|_{L_2(\mathbb{T})} \leq 1\}.$$

Consider the problem of optimal recovery of the m th derivative, $0 \leq m < r$, for functions from the class $W_2^r(\mathbb{T})$ by their inaccurate Fourier coefficients.

Let $A \subset \mathbb{Z}_+$ and $B \subset \mathbb{N}$ be finite sets (one of them may be empty). Assume that for every function $x(\cdot) \in W_2^r(\mathbb{T})$ instead of exact values of the Fourier

coefficients $a_k, k \in A$, and $b_k, k \in B$, we know their approximate values $\{\widehat{a}_k\}_{k \in A}$ and $\{\widehat{b}_k\}_{k \in B}$ such that

$$|a_k - \widehat{a}_k| \leq \delta, \quad k \in A, \quad |b_k - \widehat{b}_k| \leq \delta, \quad k \in B.$$

Denote by l_∞^s the space \mathbb{R}^s with the norm $\|\mathbf{y}\|_\infty = \max_{0 \leq j \leq s-1} |y_j|$, where $\mathbf{y} = (y_0, y_1, \dots, y_{s-1})$. Let

$$A = \{k_0, k_1, \dots, k_{N_1}\}, \quad B = \{l_1, \dots, l_{N_2}\},$$

where $k_0 < \dots < k_{N_1}$, $l_1 < \dots < l_{N_2}$ and $\text{card } A + \text{card } B = N_1 + 1 + N_2 = s$. Put

$$F_{A,B}x(\cdot) = (a_{k_0}, a_{k_1}, \dots, a_{k_{N_1}}, b_{l_1}, \dots, b_{l_{N_2}}).$$

Then we may say that we know the vector $\mathbf{y} \in l_\infty^s$ such that

$$\|F_{A,B}x(\cdot) - \mathbf{y}\|_\infty \leq \delta.$$

In this case the error of a method $\varphi: \mathbb{R}^s \rightarrow L_2(\mathbb{T})$ is defined as follows

$$e(D^m, W_2^r(\mathbb{T}), F_{A,B}, \delta, \varphi) = \sup_{\substack{x(\cdot) \in W_2^r(\mathbb{T}), \mathbf{y} \in l_\infty^s \\ \|F_{A,B}x(\cdot) - \mathbf{y}\|_\infty \leq \delta}} \|x^{(m)}(\cdot) - \varphi(\mathbf{y})(\cdot)\|_{L_2(\mathbb{T})},$$

where $D^m x(\cdot) = x^{(m)}(\cdot)$. The quantity

$$e(D^m, W_2^r(\mathbb{T}), F_{A,B}, \delta) = \inf_{\varphi: l_\infty^s \rightarrow L_2(\mathbb{T})} e(D^m, W_2^r(\mathbb{T}), F_{A,B}, \delta, \varphi)$$

is called the error of optimal recovery, and any method $\widehat{\varphi}$ for which this infimum is attained is called an optimal method of recovery.

We are interested in the problem of choosing the parameter $\widehat{\mathbf{p}} = (\widehat{A}, \widehat{B})$, $\text{card } \widehat{A} + \text{card } \widehat{B} \leq N$, so that the error of optimal recovery is minimal. In this problem we would like to consider both non-adaptive methods and adaptive ones. In terms of the general setting here $X = \mathcal{W}_2^r(\mathbb{T})$, $W = W_2^r(\mathbb{T})$, $\mathcal{Y} = \mathbb{R}^N$, $T = D^m$, $Z = L_2(\mathbb{T})$,

$$\begin{aligned} \mathcal{P} &= \mathcal{P}_N = \{(A, B) \in \mathbb{Z}_+ \times \mathbb{N} : \text{card } A + \text{card } B \leq N\}, \\ F_{\mathbf{p}}x(\cdot) &= \{\mathbf{y} \in l_\infty^s : \|F_{A,B}x(\cdot) - \mathbf{y}\|_\infty \leq \delta\}, \quad s = \text{card } A + \text{card } B. \end{aligned}$$

In [7] it was proven that for all $\mathbf{p} = (A, B) \in \mathcal{P}_N$

$$e(D^m, \mathcal{F}, \mathbf{p}) = e(D^m, W_2^r(\mathbb{T}), F_{A,B}, \delta) = \sup_{\substack{x(\cdot) \in W_2^r(\mathbb{T}) \\ \|F_{A,B}x(\cdot)\|_\infty \leq \delta}} \|x^{(m)}(\cdot)\|_{L_2(\mathbb{T})}.$$

Thus, any adaptive method (for example, those that choose the number of each next Fourier coefficient based on the values of the previously calculated coefficients) do not provide an advantage over the optimal non-adaptive method.

By virtue of the above, Theorem 4 of [7] implies the following result.

Theorem 4. *Set*

$$\chi_m = \begin{cases} 1, & m = 0, \\ 0, & m \in \mathbb{N}, \end{cases}, \quad \widehat{s} = \max \left\{ s \in \mathbb{Z}_+ : 2\delta^2 \sum_{k=0}^s k^{2r} < 1, s \leq [(N - \chi_m)/2] \right\}$$

($[a]$ is the integer part of a). Then

$$\begin{aligned} E^a(D^m, \mathcal{F}) &= E(D^m, \mathcal{F}) \\ &= \sqrt{\frac{1}{(\widehat{s} + 1)^{2(r-m)} + \frac{\delta^2}{2}\chi_m + 2\delta^2 \sum_{k=1}^{\widehat{s}} k^{2m} \left(1 - \left(\frac{k}{\widehat{s} + 1}\right)^{2(r-m)}\right)}, \end{aligned}$$

the parameter $\widehat{\mathbf{p}} = (\widehat{A}, \widehat{B})$, where for $m = 0$

$$\widehat{A} = (0, 1, \dots, \widehat{s}), \quad \widehat{B} = (1, \dots, \widehat{s}),$$

and for $m > 0$

$$\widehat{A} = (1, \dots, \widehat{s}), \quad \widehat{B} = (1, \dots, \widehat{s}).$$

The method

$$\begin{aligned} \widehat{\varphi}(\mathbf{y})(t) &= \frac{y_0}{2} \chi_m \\ &+ \sum_{k=1}^{\widehat{s}} \left(1 - \left(\frac{k}{\widehat{s} + 1}\right)^{2(r-m)}\right) k^m (y_k \cos(kt + \pi m/2) + y_{k+\widehat{s}} \sin(kt + \pi m/2)) \end{aligned}$$

is optimal.

4. Criterion for existence of a linear method which is optimal for adaptive and nonadaptive recovery

Define the mapping G_{Φ} by its graph

$$\text{gr } G_{\Phi} = \{(x, \mathbf{y}) \in W \times \mathcal{Y} : (x, \mathbf{y}) \in \text{gr } F_{\Phi(\mathbf{y})}\}.$$

Let $\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{gr } G_{\Phi}$. The value

$$r(T, \mathcal{F}, \Phi, \mathbf{y}) = \inf_{z \in Z} \sup_{x \in G_{\Phi}^{-1}(\mathbf{y})} \|Tx - z\|_Z$$

is called the Chebyshev radius of the set $T(G_{\Phi}^{-1}(\mathbf{y}))$. It is the radius of the minimal ball containing the given set. If there exists such $z(\mathbf{y}) \in Z$ that

$$r(T, \mathcal{F}, \Phi, \mathbf{y}) = \sup_{x \in G_{\Phi}^{-1}(\mathbf{y})} \|Tx - z(\mathbf{y})\|_Z,$$

then $z(\mathbf{y})$ is called the Chebyshev center of the set $T(G_{\Phi}^{-1}(\mathbf{y}))$.

The value

$$R^a(T, \mathcal{F}) = \inf_{\Phi: \mathcal{Y} \rightarrow \mathcal{P}} \sup_{\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}} r(T, \mathcal{F}, \Phi, \mathbf{y})$$

we call the adaptive radius of information in the problem of optimal adaptive recovery.

The analogous value was considered for nonadaptive recovery in [8]. It is defined as follows

$$R(T, \mathcal{F}) = \inf_{\mathbf{p} \in \mathcal{P}} \sup_{\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } F_{\mathbf{p}}} \inf_{z \in Z} \sup_{x \in F_{\mathbf{p}}^{-1}(\mathbf{y})} \|Tx - z\|_Z.$$

It follows from [8] (see also [6]) that

$$E(T, \mathcal{F}) = R(T, \mathcal{F}). \quad (7)$$

Using the same scheme of proof we obtain the analogous result for adaptive recovery.

Lemma 2. *The following equality*

$$E^a(T, \mathcal{F}) = R^a(T, \mathcal{F})$$

holds.

Proof. Let $\varphi: \mathcal{Y} \rightarrow Z$ be an arbitrary method of recovery. Then for all $\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}$

$$r(T, \mathcal{F}, \Phi, \mathbf{y}) \leq \sup_{x \in G_{\Phi}^{-1}(\mathbf{y})} \|Tx - \varphi(\mathbf{y})\|_Z \leq e^a(T, \mathcal{F}, \Phi, \varphi).$$

Taking the upper bound in the left-hand side over all $\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}$ and then taking the lower bound over all methods, we obtain

$$\sup_{\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}} r(T, \mathcal{F}, \Phi, \mathbf{y}) \leq \inf_{\varphi: \mathcal{Y} \rightarrow Z} e^a(T, \mathcal{F}, \Phi, \varphi).$$

Hence

$$R^a(T, \mathcal{F}) \leq E^a(T, \mathcal{F}). \quad (8)$$

Let us prove the opposite inequality. Let $\varepsilon > 0$. For any $\mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}$ there exists a $z_{\varepsilon}(\mathbf{y}) \in Z$ such that

$$\sup_{x \in G_{\Phi}^{-1}(\mathbf{y})} \|Tx - z_{\varepsilon}(\mathbf{y})\|_Z \leq r(T, \mathcal{F}, \Phi, \mathbf{y}) + \varepsilon.$$

We define the method $\varphi_{\varepsilon}: \mathcal{Y} \rightarrow Z$ as follows

$$\varphi_{\varepsilon}(\mathbf{y}) = \begin{cases} z_{\varepsilon}(\mathbf{y}), & \mathbf{y} \in \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}, \\ 0, & \mathbf{y} \notin \text{Pr}_{\mathcal{Y}} \text{ gr } G_{\Phi}. \end{cases}$$

Then

$$\begin{aligned} e^a(T, \mathcal{F}, \Phi, \varphi_\varepsilon) &= \sup_{(x, \mathbf{y}) \in G_\Phi} \|Tx - \varphi_\varepsilon(\mathbf{y})\|_Z \\ &= \sup_{\mathbf{y} \in \text{Pr}_Y \text{ gr } G_\Phi} \sup_{x \in G_\Phi^{-1}(\mathbf{y})} \|Tx - \varphi_\varepsilon(\mathbf{y})\|_Z \leq \sup_{\mathbf{y} \in \text{Pr}_Y \text{ gr } G_\Phi} (r(T, \mathcal{F}, \Phi, \mathbf{y}) + \varepsilon). \end{aligned}$$

Consequently,

$$E^a(T, \mathcal{F}) \leq R^a(T, \mathcal{F}) + \varepsilon.$$

In view of arbitrariness of $\varepsilon > 0$ we obtain

$$E^a(T, \mathcal{F}) \leq R^a(T, \mathcal{F}).$$

Together with (8) this proves the statement of lemma. \square

Denote by $\text{bco } A$ the convex balanced hull of A . It consists of all elements of the form

$$x = \lambda_1 x_1 + \dots + \lambda_m x_m, \quad x_j \in A, \quad j = 1, \dots, m, \quad \sum_{j=1}^m |\lambda_j| \leq 1.$$

We define the mapping $\text{bco } F_{\mathbf{p}}$ by its graph as follows

$$\text{gr bco } F_{\mathbf{p}} = \text{bco gr } F_{\mathbf{p}}.$$

Denote by \mathcal{F}^{bco} the set of all mappings $\text{bco } F_{\mathbf{p}}$.

The idea of obtaining the following result is taking from [6].

Theorem 5. *Let x' be a linear functional on X . For the existence of linear nonadaptive method with the optimal parameter $\hat{\mathbf{p}} \in \mathcal{P}$ which is optimal over all adaptive methods it is necessary and sufficient that*

$$R^a(x', \mathcal{F}) = R^a(x', \mathcal{F}^{\text{bco}}) \tag{9}$$

and

$$D(\hat{\mathbf{p}}) = \inf_{\mathbf{p} \in \mathcal{P}} D(\mathbf{p}). \tag{10}$$

Proof. Necessity. Let the method $\hat{\varphi}(\mathbf{y}) = \langle \mathbf{y}', \mathbf{y} \rangle$, where \mathbf{y}' is a linear functional on Y , and the parameter $\hat{\mathbf{p}} \in \mathcal{P}$ be such that

$$e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = E^a(x', \mathcal{F}).$$

Assume that $(x, \mathbf{y}) \in \text{bco gr } F_{\hat{\mathbf{p}}}$. Then

$$(x, \mathbf{y}) = \sum_{j=1}^m \lambda_j (x_j, \mathbf{y}_j),$$

where $(x_j, \mathbf{y}_j) \in \text{gr } F_{\hat{\mathbf{p}}}$, $j = 1, \dots, m$, and $\sum_{j=1}^m |\lambda_j| \leq 1$. Consequently,

$$\begin{aligned} |\langle x', x \rangle - \langle \mathbf{y}', \mathbf{y} \rangle| &= \left| \sum_{j=1}^m \lambda_j (\langle x', x_j \rangle - \langle \mathbf{y}', \mathbf{y}_j \rangle) \right| \\ &\leq \max_{1 \leq j \leq m} |\langle x', x_j \rangle - \langle \mathbf{y}', \mathbf{y}_j \rangle| \leq e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}). \end{aligned}$$

This implies that

$$e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) \leq e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}).$$

Since

$$\text{gr } F_{\hat{\mathbf{p}}} \subset \text{gr } \text{bco } F_{\hat{\mathbf{p}}}, \quad (11)$$

we have

$$e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) \geq e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}).$$

Thus we obtain

$$e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}). \quad (12)$$

Hence,

$$E(x', \text{bco } \mathcal{F}) \leq e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = E^a(x', \mathcal{F}).$$

In view of (4) we obtain

$$E^a(x', \text{bco } \mathcal{F}) \leq E^a(x', \mathcal{F}). \quad (13)$$

It follows from (11) that $E^a(x', \text{bco } \mathcal{F}) \geq E^a(x', \mathcal{F})$. This with (13) implies that

$$E^a(x', \text{bco } \mathcal{F}) = E^a(x', \mathcal{F}). \quad (14)$$

By virtue of Lemma 2 we obtain (9).

Now we prove (10). As it was noted above for linear functionals for any $\mathbf{p} \in \mathcal{P}$ the following equality

$$e(x', \text{bco } \mathcal{F}, \mathbf{p}) = D(\mathbf{p}) \quad (15)$$

holds. Taking into account (12) and (14) we have

$$\begin{aligned} D(\hat{\mathbf{p}}) &= e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}) \leq e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = e(x', \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) \\ &= E^a(x', \mathcal{F}) = E^a(x', \text{bco } \mathcal{F}) \leq E(x', \text{bco } \mathcal{F}) = \inf_{\mathbf{p} \in \mathcal{P}} D(\mathbf{p}). \end{aligned}$$

Sufficiency. Assume that (9) and (10) hold. In view of the fact that the set $\text{bco } \text{gr } F_{\hat{\mathbf{p}}}$ is convex and balanced, it follows from [6] that for any $\mathbf{p} \in \mathcal{P}$ there exists a $\hat{\varphi}_{\mathbf{p}}(\mathbf{y}) = \langle \mathbf{y}'_{\mathbf{p}}, \mathbf{y} \rangle$ such that

$$e(x', \text{bco } \mathcal{F}, \mathbf{p}) = e(x', \text{bco } \mathcal{F}, \mathbf{p}, \hat{\varphi}_{\mathbf{p}}).$$

Set $\hat{\varphi} = \hat{\varphi}_{\hat{\mathbf{p}}}$. Then from (15) and (10) we obtain

$$e(x', \text{bco } \mathcal{F}, \hat{\mathbf{p}}, \hat{\varphi}) = E(x', \text{bco } \mathcal{F}).$$

Moreover, it follows from (15) and Theorem 3 that

$$E(x', \text{bco } \mathcal{F}) = E^a(x', \text{bco } \mathcal{F}).$$

Therefore, taking into account (12), we obtain

$$\begin{aligned} e(x', \mathcal{F}, \widehat{\mathbf{p}}, \widehat{\varphi}) &= e(x', \text{bco } \mathcal{F}, \widehat{\mathbf{p}}, \widehat{\varphi}) = E^a(x', \text{bco } \mathcal{F}) \\ &= R^a(x', \text{bco } \mathcal{F}) = R^a(x', \mathcal{F}) = E^a(x', \mathcal{F}). \end{aligned}$$

□

The analog of Theorem 5 can not be valid for linear operators since equality (5) does not imply the existence of linear optimal method even for Hilbert spaces and exact information. We give one simple example which is the modified one from [11]. Let $X = \mathbb{R}^3$ with the Euclidean norm. Put

$$W = \{x \in \mathbb{R}^3 : |x_1| + 2|x_2| \leq 1, |x_3| \leq 2/15\}.$$

It is easy to see that W is a convex and centrally-symmetric set. We consider optimal recovery of linear operator $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ defined as follows

$$Tx = (x_1, x_2), \quad x = (x_1, x_2, x_3),$$

by exact values of linear functional $Fx = x_1 + x_3$ (here we will not mark the dependence on parameters since we have only one parameter). The norm in \mathbb{R}^2 is also Euclidean. We have

$$e(T, F) \geq D = \sup_{\substack{(x_1, x_2, x_3) \in W \\ x_1 + x_3 = 0}} \|(x_1, x_2)\| = \sup_{|x_1| + 2|x_2| \leq 1, |x_1| \leq 2/15} \|(x_1, x_2)\| = \frac{1}{2}.$$

Consider the method

$$\varphi_0(\mathbf{y}) = \begin{cases} 0, & |\mathbf{y}| \leq 4/15, \\ (\mathbf{y}, 0), & |\mathbf{y}| > 4/15. \end{cases}$$

If $|x_1 + x_3| \leq 4/15$, then

$$|x_1| = |x_1 + x_3 - x_3| \leq |x_1 + x_3| + |x_3| \leq \frac{2}{5}.$$

Consequently,

$$\sup_{\substack{(x_1, x_2, x_3) \in W \\ |x_1 + x_3| \leq 4/15}} \|(x_1, x_2) - \varphi_0(x_1 + x_3)\| = \sup_{\substack{(x_1, x_2, x_3) \in W \\ |x_1| \leq 2/5}} \|(x_1, x_2)\| = \frac{1}{2}.$$

If $(x_1, x_2, x_3) \in W$ and $|x_1 + x_3| > 4/15$, then $|x_1| \geq |x_1 + x_3| - |x_3| > 2/15$. Consequently, $|x_2| < 13/30$. Therefore

$$\begin{aligned} \sup_{\substack{(x_1, x_2, x_3) \in W \\ |x_1 + x_3| > 4/15}} \|(x_1, x_2) - \varphi_0(x_1 + x_3)\| &\leq \sup_{\substack{(x_1, x_2, x_3) \in W \\ |x_2| < 13/30}} \|(-x_3, x_2)\| \\ &< \sqrt{\frac{4}{225} + \frac{169}{900}} < \frac{1}{2}. \end{aligned}$$

Thus,

$$e(T, F) = D = \frac{1}{2}.$$

Now we show that for any linear method the error of recovery is greater than $1/2$. For any linear method $\varphi(\mathbf{y}) = (c_1\mathbf{y}, c_2\mathbf{y})$, $c_1, c_2 \in \mathbb{R}$, we have

$$e(T, F, \varphi) = \sup_{(x_1, x_2, x_3) \in W} \sqrt{(x_1 - c_1(x_1 + x_3))^2 + (x_2 - c_2(x_1 + x_3))^2}.$$

If $c_1 \leq 0$, considering the point $(1, 0, 0) \in W$, we obtain

$$e(T, F, \varphi) \geq \sqrt{(1 - c_1)^2 + c_2^2} \geq 1.$$

If $c_1 > 0$, considering the point $(0, 1/2, 2/15 \operatorname{sign} c_2) \in W$, we obtain

$$e(T, F, \varphi) \geq \sqrt{c_1^2 \frac{4}{15} + \left(\frac{1}{2} + |c_2| \frac{2}{15}\right)^2} > \frac{1}{2}.$$

Consequently,

$$e(T, F, \varphi) > \frac{1}{2} = e(T, F).$$

In the last example we see that there is no linear optimal method of recovery but condition (5) is satisfied. We give one more example to show that in general this condition is not fulfilled even for Hilbert spaces and exact information.

Let $X = \mathbb{R}^3$ with the Euclidean norm $Tx = x$ and $Fx = x_1$, $x = (x_1, x_2, x_3)$. Consider the equilateral triangle Δ on the plane $x_1 = 1$ with vertexes at the points $A(1, 0, 1/\sqrt{3})$, $B(1, -1/2, -1/(2\sqrt{3}))$, $C(1, 1/2, -1/(2\sqrt{3}))$. Put $W = \operatorname{bco} \Delta$ (in other words, W is the convex hull of Δ and $-\Delta$). It may be easily shown that the intersection of W with the plane $x_1 = 0$ is the equilateral hexagon with the center at origin and with a side equal to $1/2$. Thus,

$$D = \sup_{\substack{x \in W \\ Fx=0}} \|Tx\| = \sup_{\substack{x \in W \\ x_1=0}} \|x\| = \frac{1}{2}.$$

On the other hand, in view of (7) (here we do not mark the dependence on parameters again since we have only one parameter) we have

$$e(T, F) \geq \inf_{z \in \mathbb{R}^3} \sup_{\substack{x \in W \\ x_1=1}} \|x - z\|_Z = \frac{1}{\sqrt{3}} > \frac{1}{2}.$$

Consequently, $e(T, F) > D$.

The author is very grateful to the referees for a number of helpful comments and suggestions.

References

- [1] N.S. Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions, *USSR Comput. Math. Math. Phys.* 11 (1971) 244–249.
- [2] S. Gal, C.A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, *Appl. Anal.* 10 (2) (1980) 105–120.
- [3] M.A. Kon, E. Novak, On the adaptive and continuous information problems, *J. Complexity* 5 (1989) 345–362.
- [4] M.A. Kon, E. Novak, The adaption problem for approximating linear operators, *Bull. Amer. Math. Soc.* 23 (1990) 159–165.
- [5] N.P. Korneichuk, Optimization of active algorithms for recovery of monotonic functions from Hölder’s class, *J. Complexity* 10 (1994) 265–269.
- [6] G.G. Magaril-II’yaev, K.Yu. Osipenko, Optimal recovery of functionals based on inaccurate data, *Math. Notes* 50 (6) (1991) 1274–1279.
- [7] G.G. Magaril-II’yaev, K.Yu. Osipenko, On best harmonic synthesis of periodic functions, *J. Math. Sci.* 209 (1) (2015) 115–129.
- [8] C.A. Micchelli, T.J. Rivlin, A survey of optimal recovery, *Optimal estimation in approximation theory*, Proc. Internat. Sympos. (Freudenstadt 1976), Plenum, New York 1977, 1–54.
- [9] E. Novak, Quadrature formulas for convex classes of functions, Ed. H. Braß and G. Hämmerlin, *Numerical Integration IV*, Birkhäuser Verlag, Basel, 1993, 283–296.
- [10] E. Novak, The adaption problem for nonsymmetric convex sets, *J. Approx. Theory* 82 (1) (1995) 123–134.
- [11] K.Yu. Osipenko, Optimal recovery of linear functionals and operators, *Communication on Applied Mathematics and Computation* 30 (4) (2016) 459–482.
- [12] L. Plaskota, A note on varying cardinality in the average case setting, *J. Complexity* 9 (1993) 458–470.
- [13] L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge University Press, 1996.
- [14] J.F. Traub, G.W. Wasilkowski, and H. Woźniakowski, *Information-based Complexity*, Academic Press, New York, 1988.
- [15] J.F. Traub, H. Woźniakowski, *A General Theory of Optimal Algorithms*, New York, Academic Press, 1980.
- [16] G.W. Wasilkowski, Information of varying cardinality, *J. Complexity* 2 (1986) 204–228.