数理解析研究所講究録
第 1980 巻 2016 年 1-15

Adaptive Learning with Reproducing Kernels

Masahiro Yukawa
Department of Electronics and Electrical Engineering, Keio University

1 An Overview

Reproducing kernels have been proven an attractive tool in the context of online estimation of nonlinear functions over the last decades in the signal processing and machine learning communities [1–16]. (See [17–27] for the theory and applications of reproducing kernels.) The approach of online nonlinear estimation with kernels has mild computational complexity compared to the approach based on Volterra series expansion (of which the second or third order approximation is typically used) and has convex nature of stochastic optimization unlike the neural network. The challenges of the kernel-based approach include

a) kernel design (how to design a reproducing kernel that fits the nonlinear function to be estimated);

b) dictionary construction (how to construct a *dictionary*, a set of vectors, that spans a 'low-dimensional' linear subspace containing a vector close to the nonlinear function); and

c) parameter estimation (how to estimate the coefficients of the dictionary elements to approximate the nonlinear function in online and adaptive fashion).

Here, the low dimensionality is of great importance in practice for efficiency both in computation and memory-resource. For item c), any algorithmic solvers that have been developed for linear adaptive filtering tasks can be directly applied, such as the adaptive projected subgradient method (APSM) [28–43], the adaptive proximal forward-backward splitting (APFBS) method [44–48], etc. APSM is an adaptive extension of the projected subgradient method studied by Polyak [49], and it includes as its particular case the classical normalized least mean square (NLMS) algorithm [50, 51], the affine projection algorithm [52–54], the adaptive parallel subgradient projection algorithm [55–58], and their constrained versions [59–61]. It formulates the adaptive estimation tasks as an asymptotic minimization problem of a sequence of nonnegative convex functions, and a strong convergence theorem has been established in [28] under certain mild conditions. As a direct consequence of the convergence theorem, it extends the convergence theorem for the projections onto convex sets (POCS), a celebrated alternating projection method for convex feasibility problems (cf. [62]), to the case of infinitely many closed convex sets. It should be remarked here that the strong convergence is proved for APSM, whereas the widely-known result for POCS is weak convergence (see, e.g., [63]). APFBS is an adaptive extension of the proximal forward-backward splitting method [64,65] for minimizing a sum of smooth and nonsmooth convex functions by using Moreau's proximity operator [66–68]. Its typical applications include adaptive estimation of sparse impulse responses [44].

For item a), the author has proposed *multikernel adaptive filtering* [69–72], a convex-analytic learning paradigm using 'multiple' reproducing kernels; the reader may refer to the tutorial paper [73]. Multikernel adaptive filtering is particularly effective when (i) the nonlinear function contains multiple components with different characteristics such as linear and nonlinear components and high- and low- frequency components [74–77], and (ii) an adequate kernel is unavailable because the amount of prior information about the unknown function is limited, and/or because the unknown function is time-varying and so is an adequate kernel for the function. Related approaches have been considered by different research groups [78–80].

The author has also proposed an efficient single-kernel adaptive filtering algorithm named hyperplane projection along affine subspace (HYPASS) [81–83]. The HYPASS algorithm is a natural extension of the simple stochastic-gradient-based method called the naive online R_{reg} minimization algorithm (NORMA) proposed by Kivinen *et al.* [2]. NORMA builds a dictionary by using all the observed data, meaning that the dictionary size grows linearly with the number of data observed. As a remedy for this issue, a simple truncation rule has been introduced in [2]. This approach is apparently inefficient because the dictionary may contain redundant vectors, which cause high dimensionality of the subspace spanned by the dictionary. HYPASS selectively adds each observed datum into the dictionary based on the so-called coherence criterion [9]; other criteria have also been proposed, e.g., in [3, 84]. If a datum does not enter the dictionary, the stochastic-gradient direction does not belong to the dictionary subspace and thus the datum is simply discarded. In other words, as long as sticking to the stochastic gradient method, one has to discard such a datum although it may contain information for updating the coefficients. The HYPASS algorithm systematically eliminates this limitation by enforcing the update direction to lie in the dictionary subspace. HYPASS includes the sparse sequential method of Dodd *et al.* [85] and the quantized kernel LMS (QKLMS) [12] as particular cases. Another technique that has been developed by the author is the adaptive refinement of the dictionary [71, 86, 87], borrowing the idea of sparse signal recovery such as compressed sensing [88–90]. (See [91] for a sparse signal recovery using non-quadratic strictly-convex objectives. See also [92, 93] for studies of regularization paths with ℓ_p quasi-norms for $0 < p < 1$.) It has also been extended to online model selection and learning scheme in [94–96], and the scheme has successfully been applied to an adaptive online coverage-map reconstruction problem in wireless communications [97].

The existing algorithms of online nonlinear estimation with kernels can be classified into two categories: the functional approach and the parameter approach. Here, the functional approach formulates the online nonlinear estimation problem in a reproducing kernel Hilbert space (RKHS), while the parameter approach formulates the problem in a parameter (Euclidean) space. Our recent studies have shown significant advantages of the functional approach over the parameter space [81–83]. The Cartesian HYPASS (CHYPASS) algorithm proposed in [72] is a multikernel adaptive filtering scheme falling into the functional approach, which unifies the works in [71] and [81–83]. CHYPASS has been applied to time-series prediction problems with laser signals and CO2 emission data, and its efficacy has been demonstrated in [72]. In the remainder of this article, we present basic materials that support the CHYPASS algorithm, and then present its concept in a simple manner.

2 Sum Space and Reproducing Kernel

We denote by $\mathbb{R}$ and $\mathbb{N}$ the sets of all real numbers and nonnegative integers, respectively. Vectors and matrices are denoted by lower-case and upper-case letters in bold-face, respectively. The identity matrix is denoted by $\boldsymbol{I}$ and the transposition of a vector/matrix is denoted by $(\cdot)^{\mathsf{T}}$. We denote the null (zero) function by 0.

Let $\mathcal{U} \subset \mathbb{R}^L$ and $\mathbb{R}$ be the input and output spaces, respectively. We consider the problem of estimating/tracking a nonlinear unknown function $\psi : \mathcal{U} \to \mathbb{R}$ from sequentially arriving input-output measurements $\boldsymbol{u}_n \in \mathcal{U}$ and $d_n := \psi(\boldsymbol{u}_n) + v_n \in \mathbb{R}$, $n \in \mathbb{N}$, where $v_n \in \mathbb{R}$ is the additive noise. We focus on the case where ψ contains several distinctive components; e.g., linear and nonlinear (but smooth) components, high- and low-frequency components, etc. To generate a minimal model to describe such a multicomponent function ψ, we use multiple RKHSs $(\mathcal{H}_1, \langle\cdot,\cdot\rangle_{\mathcal{H}_1})$, $(\mathcal{H}_2, \langle\cdot,\cdot\rangle_{\mathcal{H}_2})$, $\cdots$, $(\mathcal{H}_Q, \langle\cdot,\cdot\rangle_{\mathcal{H}_Q})$ over $\mathcal{U}$, where each of the $\mathcal{H}_q$s consists of functions from $\mathcal{U}$ to $\mathbb{R}$. Here, Q is the number of components of ψ, and each RKHS is associated with each component. The positive definite kernel associated with the qth RKHS $\mathcal{H}_q$, $q \in \mathcal{Q} := \{1, 2, \cdots, Q\}$, is denoted by $\kappa_q : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$, and the norm induced by $\langle\cdot,\cdot\rangle_{\mathcal{H}_q}$ is denote by $\|\cdot\|_{\mathcal{H}_q}$. The function ψ is modeled as an element of the sum space

$$\mathcal{H}^+ := \mathcal{H}_1 + \mathcal{H}_2 + \cdots + \mathcal{H}_Q := \left\{ \sum_{q \in \mathcal{Q}} f_q : f_q \in \mathcal{H}_q \right\}.$$

Given an element $f \in \mathcal{H}^+$, the decomposition $f = \sum_{q \in \mathcal{Q}} f_q$, $f_q \in \mathcal{H}_q$, is not necessarily unique in general. If the decomposition is unique for any $f \in \mathcal{H}^+$, $\mathcal{H}^+$ is the *direct sum* of $\mathcal{H}_q$s [98], and it is usually denoted by $\mathcal{H}^+ = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots \oplus \mathcal{H}_Q$.

Theorem 1 (Reproducing kernel of sum space $\mathcal{H}^+$ [17]) *The sum space $\mathcal{H}^+$ equipped with the norm*

$$\|f\|_{\mathcal{H}^+}^2 := \min \left\{ \sum_{q \in \mathcal{Q}} \|f_q\|_{\mathcal{H}_q}^2 \mid f = \sum_{q \in \mathcal{Q}} f_q, \ f_q \in \mathcal{H}_q \right\}, \quad f \in \mathcal{H}^+, \tag{1}$$

is a RKHS with the reproducing kernel $\kappa := \sum_{q \in \mathcal{Q}} \kappa_q$.

Theorem 2 *Let $\kappa : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ be the reproducing kernel of a real Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$. Then, given an arbitrary $w > 0$, $\kappa_w(\boldsymbol{u}, \boldsymbol{v}) := w\kappa(\boldsymbol{u}, \boldsymbol{v})$, $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{U}$, is the reproducing kernel of the RKHS $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H},w})$ with the inner product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathcal{H},w} := w^{-1} \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathcal{H}}$, $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{U}$.*

Theorems 1 and 2 yield the following result.

Corollary 1 (Weighted norm and reproducing kernel) *Given any $w_q > 0$, $q \in \mathcal{Q}$, $\kappa_{\boldsymbol{w}}(\boldsymbol{u}, \boldsymbol{v}) := \sum_{q \in \mathcal{Q}} w_q \kappa_q(\boldsymbol{u}, \boldsymbol{v})$, $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{U}$, is the reproducing kernel of the sum space $\mathcal{H}^+$ equipped with the weighted norm $\|\cdot\|_{\mathcal{H}^+,\boldsymbol{w}}$ defined as $\|f\|_{\mathcal{H}^+,\boldsymbol{w}}^2 := \min \left\{ \sum_{q \in \mathcal{Q}} w_q^{-1} \|f_q\|_{\mathcal{H}_q}^2 \mid f = \sum_{q \in \mathcal{Q}} f_q, \ f_q \in \mathcal{H}_q \right\}$, $f \in \mathcal{H}^+$.*

3 Examples of Reproducing Kernels and Basic Results

Example 1 (Positive definite kernels)

1. *Linear kernel: Given $c \geq 0$,*

$$\kappa_{\mathrm{L}}(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y} + c, \ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U}. \tag{2}$$

2. *Polynomial kernel: Given $c \geq 0$ and $m \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}$,*

$$\kappa_{\mathrm{P}}(\boldsymbol{x}, \boldsymbol{y}) := (\boldsymbol{x}^{\mathsf{T}} \boldsymbol{y} + c)^m, \ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U}. \tag{3}$$

3. *Gaussian kernel (normalized): Given $\sigma > 0$,*

$$\kappa_{\mathrm{G},\sigma}(\boldsymbol{x}, \boldsymbol{y}) := \frac{1}{(\sqrt{2\pi}\sigma)^L} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^L}^2}{2\sigma^2} \right), \ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U}. \tag{4}$$

The following theorem has been shown by Minh in 2010.

Theorem 3 ([99]). *Let $\mathcal{U} \subset \mathbb{R}^L$ be any set with nonempty interior and $\mathcal{H}_{\kappa_{G,\sigma}}$ the RKHS associated with a Gaussian kernel $\kappa_{G,\sigma}(\boldsymbol{x}, \boldsymbol{y})$ for an arbitrary $\sigma > 0$ together with the input space $\mathcal{U}$. Then, $\mathcal{H}_{\kappa_{G,\sigma}}$ does not contain any polynomial on $\mathcal{U}$, including the nonzero constant function.*

Corollary 2 (Polynomial and Gaussian RKHSs [72]) *Assume that the input space $\mathcal{U} \subset \mathbb{R}^L$ has nonempty interior. Let $\mathcal{H}_{\kappa_{P,\sigma}}$ and $\mathcal{H}_{\kappa_{G,\sigma}}$ be the RKHSs associated, respectively, with a polynomial kernel κ_P and a Gaussian kernel $\kappa_{G,\sigma}$ for arbitrary parameters $c \geq 0$, $m \in \mathbb{N}^*$, and $\sigma > 0$. Then,*

$$\mathcal{H}_{\kappa_P} \cap \mathcal{H}_{\kappa_{G,\sigma}} = \{0\}. \tag{5}$$

In particular, (5) for $m = 1$ implies that

$$\mathcal{H}_{\kappa_L} \cap \mathcal{H}_{\kappa_{G,\sigma}} = \{0\}. \tag{6}$$

Theorem 4 ([72]) *Let $\mathcal{U} \subset \mathbb{R}^L$ be an arbitrary subset and $\kappa_1 := w_1\kappa_{G,\sigma_1}$ and $\kappa_2 := w_2\kappa_{G,\sigma_2}$ Gaussian kernels for $\sigma_1 > \sigma_2 > 0$ and $w_1, w_2 > 0$. Then, the associated RKHSs $\mathcal{H}_1$ and $\mathcal{H}_2$ satisfy the following:*

1. *$\mathcal{H}_1 \subset \mathcal{H}_2$, and*
2. *$\sqrt{w_1}\,\|f\|_{\mathcal{H}_1} \geq \sqrt{w_2}\,\|f\|_{\mathcal{H}_2}$ for any $f \in \mathcal{H}_1$.*

See [100–102] for the related results to Theorem 4.

4 Adaptive Leaning Algorithm Based on Orthogonal Projection in Product Space

Suppose that $\mathcal{H}_p \cap \mathcal{H}_q = \{0\}$ for any $p \neq q$ (cf. Corollary 2). Then, any $f \in \mathcal{H}^+$ can be *decomposed uniquely* into $f = \sum_{q\in\mathcal{Q}} f_q$, $f_q \in \mathcal{H}_q$. It is clear in this case that, under the correspondence between f and the Q-tuple $(f_q)_{q\in\mathcal{Q}}$, the sum space $\mathcal{H}^+$ is isomorphic to the Cartesian product

$$\mathcal{H}^\times := \mathcal{H}_1 \times \mathcal{H}_2 \times \cdots \times \mathcal{H}_Q := \{(f_1, f_2, \cdots, f_Q) : f_q \in \mathcal{H}_q,\ q \in \mathcal{Q}\},$$

which is a real Hilbert space equipped with the inner product defined by

$$\langle f, g\rangle_{\mathcal{H}^\times} := \sum_{q\in\mathcal{Q}} \langle f_q, g_q\rangle_{\mathcal{H}_q}, \quad f = (f_q)_{q\in\mathcal{Q}},\ g = (g_q)_{q\in\mathcal{Q}} \in \mathcal{H}^\times. \tag{7}$$

We denote by $\mathcal{D}_{q,n} \subset \{\kappa_q(\cdot, \boldsymbol{u}) \mid \boldsymbol{u} \in \mathcal{U}\}$ the *dictionary* constructed for the qth kernel at time $n \in \mathbb{N}$. The dictionary typically starts with $\mathcal{D}_{q,-1} := \emptyset$ and grows based on some novelty criterion such as the coherence criterion (dictionary constructions for CHYPASS depend on kernels employed; see [72] for details). The *kernel-by-kernel dictionary subspaces* are defined as $\mathcal{M}_{q,n} := \operatorname{span} \mathcal{D}_{q,n} \subset \mathcal{H}_q$, $q \in \mathcal{Q}$, $n \in \mathbb{N}$. The multikernel adaptive filter at time n is given as

$$\varphi_n := \sum_{q \in \mathcal{Q}} \varphi_{q,n} \in \mathcal{H}^+, \ n \in \mathbb{N}, \tag{8}$$

where $\varphi_{q,n} \in \mathcal{M}_{q,n-1}$. Thus, the dictionary $\mathcal{D}_{q,n}$ contains the atoms (vectors) that form the next estimate $\varphi_{q,n+1}$.

It should be emphasized that the isomorphism mentioned above relies on the assumption $\mathcal{H}_p \cap \mathcal{H}_q = \{0\}$, $\forall p \neq q$. In general, the norm in the sum space has no closed-form, as can be seen from Corollary 1. This makes the orthogonal projection in the sum space difficult to compute in practice. The CHYPASS algorithm therefore projects the current estimate onto a zero instantaneous-error hyperplane in the product space (rather than in the sum space). After simple manipulations, with the initial filter $\varphi_0 := 0$, its update equation is given as follows [72]:

$$\varphi_{n+1} := \varphi_n + \lambda_n \frac{d_n - \varphi_n(\boldsymbol{u}_n)}{\sum_{q \in \mathcal{Q}} \left\| P_{\mathcal{M}_{q,n}}(\kappa_q(\cdot, \boldsymbol{u}_n)) \right\|_{\mathcal{H}_q}^2} \sum_{q \in \mathcal{Q}} P_{\mathcal{M}_{q,n}}(\kappa_q(\cdot, \boldsymbol{u}_n)), \quad n \in \mathbb{N}, \tag{9}$$

where $\lambda_n \in (0, 2)$ is the step size and $P_{\mathcal{M}_{q,n}}(\kappa_q(\cdot, \boldsymbol{u}_n))$ denotes the orthogonal projection of $\kappa_q(\cdot, \boldsymbol{u}_n)$ onto the closed linear subspace $\mathcal{M}_{q,n}$ [98]. This can be computed efficiently.

When we employ a linear (or polynomial) kernel together with a Gaussian kernel, the product space $\mathcal{H}^\times$ is isomorphic to the sum space $\mathcal{H}^+$ by Corollary 2, which implies that CHYPASS can be interpreted as projecting the current estimate into the zero instantaneous-error hyperplane in the sum space $\mathcal{H}^+$ in this case. Referring to Theorem 4, on the other hand, the same does not apply to the case where we employ multiple Gaussian kernels. CHYPASS works well in both cases, as demonstrated by simulations in [72].

References

[1] L. Csato and M. Opper, *Sparse representation for Gaussian process models,* in Advances in Neural Information Processing Systems 13. MIT Press, 2001, pp. 444–450.

[2] J. Kivinen, A. J. Smola, and R. C. Williamson, "Online learning with kernels," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2165–2176, Aug. 2004.

[3] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[4] A. V. Malipatil, Y.-F. Huang, S. Andra, and K. Bennett, "Kernelized set-membership approach to nonlinear adaptive filtering," in *Proc. IEEE ICASSP*, 2005, pp. 149–152.

[5] P. Laskov, C.Gehl, S.Krüger, and K.-R. Müller, "Incremental support vector learning: Analysis, implementation and applications," *J. Mach. Learn. Res.*, vol. 7, pp. 1909–1936, 2006.

[6] W. Liu and J. Príncipe, "Kernel affine projection algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–12, 2008, article ID 784292.

[7] K. Slavakis, S. Theodoridis, and I. Yamada, "Online kernel-based classification using adaptive projection algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2781–2796, July 2008.

[8] ——, "Adaptive constrained learning in reproducing kernel Hilbert spaces: the robust beamforming case," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4744–4764, Dec. 2009.

[9] C. Richard, J. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1058–1067, Mar. 2009.

[10] W. Liu, J. Príncipe, and S. Haykin, *Kernel Adaptive Filtering.* New Jersey: Wiley, 2010.

[11] S. Van Vaerenbergh, M. Lázaro-Gredilla, and I. Santamaría, "Kernel recursive least-squares tracker for time-varying regression," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1313–1326, Aug. 2012.

[12] B. Chen, S. Zhao, P. Zhu, and J. C. Príncipe, "Quantized kernel least mean square algorithm," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 1, pp. 22–32, Jan. 2012.

[13] K. Slavakis, P. Bouboulis, and S. Theodoridis, "Online learning in reproducing kernel Hilbert spaces," in *Academic Press Library in Signal Processing: 1st Edition, Signal Processing Theory and Machine Learning.* Elsevier, 2014, vol. 1, pp. 883–987.

[14] W. Gao, J. Chen, C. Richard, and J. Huang, "Online dictionary learning for kernel LMS," *IEEE Trans. Signal Processing*, vol. 62, no. 11, pp. 2765–2777, Jun. 2014.

[15] H. Fan, Q. Song, and S. B. Shrestha, "Online learning with kernel regularized least mean square algorithms," *Knowledge-Based Systems*, vol. 59, pp. 21–32, Mar. 2014.

[16] P. Honeine, "Analyzing sparse dictionaries for online learning with kernels," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6343–6353, Dec. 2015.

[17] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337–404, May 1950.

[18] S. Saitoh, *Integral Transforms, Reproducing Kernels and Their Applications*, ser. Pitman Research Notes in Mathematics. Addison Wesley Longman, 1997, vol. 369.

[19] V. N. Vapnik, *Statistical Learning Theory.* New York: Wiley, 1998.

[20] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, no. 2, pp. 181–201, Mar. 2001.

[21] B. Schölkopf and A. J. Smola, *Learning with Kernels.* Cambridge, MA: MIT Press, 2001.

[22] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics.* Boston: MA: Kluwer Academic, 2004.

[23] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis.* Cambridge University Press, 2004.

[24] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. New York: Academic, 2008.

[25] 福水健次, **カーネル法入門**. 朝倉出版, 2010.

[26] *IEEE Signal Processing Magazine, Special Section: Advances in Kernel-based Learning for Signal Processing*, July 2013.

[27] Y. Motai, "Kernel association for classification and prediction: a survey," *IEEE Trans. Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 208–223, Feb. 2015.

[28] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numer. Funct. Anal. Optim.*, vol. 25, no. 7&8, pp. 593–617, 2004.

[29] M. Yukawa, R. L. G. Cavalcante, and I. Yamada, "Efficient blind MAI suppression in DS/CDMA systems by embedded constraint parallel projection techniques," *IEICE Trans. Fundamentals*, vol. E88-A, no. 8, pp. 2062–2071, Aug. 2005.

[30] R. L. G. Cavalcante, M. Yukawa, and I. Yamada, "Set-theoretic DS/CDMA receivers for fading channels by adaptive projected subgradient method," in *Proc. IEEE GLOBECOM*, SP06-1, 2005.

[31] K. Slavakis, I. Yamada, and N. Ogura, "The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings," *Numer. Funct. Anal. Optim.*, vol. 27, no. 7&8, pp. 905–930, 2006.

[32] M. Yukawa, K. Slavakis, and I. Yamada, "Adaptive parallel quadratic-metric projection algorithms," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 5, pp. 1665–1680, July 2007.

[33] R. L. G. Cavalcante and I. Yamada, "Multiaccess interference suppression in orthogonal space-time block coded MIMO systems by adaptive projected subgradient method," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1028–1042, Mar. 2008.

[34] ——, "A flexible peak-to-average power ratio reduction scheme for OFDM systems by the adaptive projected subgradient method," *IEEE Trans. Signal Processing*, vol. 57, no. 4, pp. 1456–1468, Apr. 2009.

[35] M. Yukawa, "Krylov-proportionate adaptive filtering techniques not limited to sparse systems," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 927–943, Mar. 2009.

[36] M. Yukawa, R. C. de Lamare, and I. Yamada, "Robust reduced rank adaptive algorithm based on parallel subgradient projection and krylov subspace," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4660–4674, Dec. 2009.

[37] M. Yukawa and I. Yamada, "A unified view of adaptive variable-metric projection algorithms," *EURASIP J. Advances in Signal Processing*, vol. 2009, Article ID 589260, 13 pages, 2009.

[38] M. Yukawa, K. Slavakis, and I. Yamada, "Multi-domain adaptive learning based on feasibility splitting and adaptive projected subgradient method," *IEICE Trans. Fundamentals*, vol. E93-A, no. 2, pp. 456–466, Feb. 2010.

[39] M. Yukawa and W. Utschick, "A fast stochastic gradient algorithm: Maximal use of sparsification benefits under computational constraints," *IEICE Trans. Fundamentals*, vol. E93-A, no. 2, pp. 467–475, Feb. 2010.

[40] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections: a unifying framework for linear and nonlinear classification and regression tasks," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 97–123, Jan. 2011.

[41] M. Yukawa and I. Yamada, "Set-theoretic adaptive filtering based on data-driven sparsification," *International Journal of Adaptive Control and Signal Processing*, vol. 25, pp. 707–722, Wiley, 2011.

[42] M. Yukawa, Y. Sung, and G. Lee, "Dual-domain adaptive beamformer under linearly and quadratically constrained minimum variance," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2874–2886, Jun. 2013.

[43] O. Toda, M. Yukawa, S. Sasaki, and H. Kikuchi, "An efficient adaptive filtering scheme based on combining multiple metrics," *IEICE Trans. Fundamentals*, vol. E97-A, no. 3, pp. 800–808, Mar. 2014.

[44] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A sparse adaptive filtering using time-varying soft-thresholding techniques," in *Proc. IEEE ICASSP*, 2010, pp. 3734–3737.

[45] M. Yamagishi, M. Yukawa, and I. Yamada, "Sparse system identification by exponentially weighted adaptive parallel projection and generalized soft-thresholding," in *Proc. APSIPA Annual Summit and Conference*, 2010, pp. 367–370.

[46] M. Yukawa, Y. Tawara, M. Yamagishi, and I. Yamada, "Sparsity-aware adaptive filters based on ℓ_p-norm inspired soft-thresholding technique," in *Proc. IEEE ISCAS*, 2012, pp. 2749–2752.

[47] M. Yukawa, Y. Tawara, S. Sasaki, and I. Yamada, "A sparsity-based design of regularization parameter for adaptive proximal forward-backward splitting algorithm," in *Proc. IEEE ISWCS*, 2013, pp. 190–193.

[48] M. Yamagishi, M. Yukawa, and I. Yamada, "Shrinkage tuning based on an unbiased MSE estimate for sparsity-aware adaptive filtering," in *Proc. IEEE ICASSP*, 2014, pp. 5514–5518.

[49] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Comput. Math. Physics 9*, pp. 14–29, 1969.

[50] J. Nagumo and J. Noda, "A learning method for system identification," *IEEE Trans. Autom. Control*, vol. 12, no. 3, pp. 282–287, 1967.

[51] A. E. Albert and L. S. Gardner Jr., *Stochastic Approximation and Nonlinear Regression.* Cambridge MA: MIT Press, 1967.

[52] T. Hinamoto and S. Maekawa, "Extended theory of learning identification," *Trans. Inst. Elect. Eng. Jpn.*, vol. 95, no. 10, pp. 227–234, 1975 (in Japanese).

[53] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.

[54] M. Rupp, "A family of adaptive filter algorithms with decorrelating properties," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 771–775, Mar. 1998.

[55] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1091–1101, May 2002.

[56] M. Yukawa and I. Yamada, "Efficient adaptive stereo echo canceling schemes based on simultaneous use of multiple state data," *IEICE Trans. Fundamentals*, vol. E87-A, no. 8, pp. 1949–1957, Aug. 2004.

[57] ——, "Pairwise optimal weight realization —Acceleration technique for set-theoretic adaptive parallel subgradient projection algorithm," *IEEE Trans. Signal Processing*, vol. 54, no. 12, pp. 4557–4571, Dec. 2006.

[58] M. Yukawa, N. Murakoshi, and I. Yamada, "Efficient fast stereo acoustic echo cancellation based on pairwise optimal weight realization technique," *EURASIP J. Appl. Signal Processing*, vol. 2006, Article ID 84797, 15 pages, 2006.

[59] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[60] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[61] M. L. R. de Campos, S. Werner, and J. A. Apolinário Jr., "Constrained adaptation algorithms employing householder transformation," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2187–2195, Sept. 2002.

[62] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. of IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[63] H. Stark and Y. Yang, *Vector Space Projections—A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* New York: John Wiley & Sons, 1998.

[64] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, 1979.

[65] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, pp. 1168–1200, 2005.

[66] J. J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *C. R. Acad. Sci. Paris Sér*, vol. 255, pp. 2897–2899, 1962.

[67] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 1st ed. New York: NY: Springer, 2011.

[68] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau envelope of nonsmooth convex functions over the fixed point set of certain quasi-nonexpansive mappings," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Optimization and Its Applications, vol. 49. New York: Springer, 2011, pp. 345–390.

[69] M. Yukawa, "On use of multiple kernels in adaptive learning —Extended reproducing kernel Hilbert space with Cartesian product," in *Proc. IEICE Signal Processing Symposium*, Nov. 2010, pp. 59–64.

[70] ——, "Nonlinear adaptive filtering techniques with multiple kernels," in *European Signal Processing Conference (EUSIPCO)*, 2011, pp. 136–140.

[71] ——, "Multikernel adaptive filtering," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4672–4682, Sept. 2012.

[72] ——, "Adaptive learning in Cartesian product of reproducing kernel Hilbert spaces," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6037–6048, Nov. 2015.

[73] 湯川正裕, "非線形適応信号処理技術の新潮流:再生核の応用," **電子情報通信学会誌**, vol. 97, no. 10, pp. 876–882, Oct. 2014.

[74] W. Härdle, H. Liang, and J. Gao, *Partially Linear Models.* Heidelberg, Germany: Physica-Verlag, 2000.

[75] M. Espinoza, J. A. K. Suykens, and B. D. Moor, "Kernel based partially linear models and nonlinear identification," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1602–1606, Oct. 2005.

[76] Y.-L. Xu and D.-R. Chen, "Partially-linear least-squares regularized regression for system identification," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2637–2641, Nov. 2009.

[77] Y.-L. Xu, D.-R. Chen, H.-X. Li, and L. Liu, "Least square regularized regression in sum space," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 635–646, Apr. 2013.

[78] R. Pokharel, J. Príncipe, and S. Seth, "Mixture kernel least mean square," in *IEEE IJCNN*, 2013.

[79] W. Gao, C. Richard, J.-C. M. Bermudez, and J. Huang, "Convex combinations of kernel adaptive filters," in *IEEE Int. Workshop on MLSP*, 2014.

[80] F. A. Tobar, S.-Y. Kung, and D. P. Mandic, "Multikernel least mean square algorithm," *IEEE Trans. Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 265–277, Feb. 2014.

[81] M. Yukawa and R. Ishii, "An efficient kernel adaptive filtering algorithm using hyperplane projection along affine subspace," in *Proc. EUSIPCO*, 2012, pp. 2183–2187.

[82] M. Takizawa and M. Yukawa, "An efficient data-reusing kernel adaptive filtering algorithm based on parallel hyperslab projection along affine subspaces," in *Proc. IEEE ICASSP*, 2013, pp. 3557–3561.

[83] ——, "Adaptive nonlinear estimation based on parallel projection along affine subspaces in reproducing kernel Hilbert space"," *IEEE Trans. Signal Processing*, vol. 63, no. 16, pp. 4257–4269, Aug. 2015.

[84] J. Platt, "A resource-allocating network for function interpolation," *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.

[85] T. J. Dodd, V. Kadirkamanathan, and R. F. Harrison, "Function estimation in Hilbert space using sequential projections," in *IFAC Conf. Intell. Control Syst. Signal Process.*, 2003, pp. 113–118.

[86] M. Takizawa and M. Yukawa, "An efficient sparse kernel adaptive filtering algorithm based on isomorphism between functional subspace and Euclidean space," in *Proc. IEEE ICASSP*, 2014, pp. 4541–4545.

[87] ——, "Efficient dictionary-refining kernel adaptive filter with fundamental insights," submitted for publication.

[88] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing.* New York: Springer, 2010.

[89] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity.* New York: Cambridge University Press, 2010.

[90] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing.* Birkhäuser, 2013.

[91] S. Amari and M. Yukawa, "Minkovskian gradient for sparse optimization," *IEEE J. Selected Topics in Signal Process.*, vol. 4, no. 4, pp. 576–585, Aug. 2013.

[92] K. Jeong, M. Yukawa, and S. Amari, "Can critical-point paths under ℓ_p-regularization $(0 < p < 1)$ reach the sparsest least squares solutions?" *IEEE Trans. Information Theory*, vol. 69, no. 5, pp. 2960–2968, May 2014.

[93] M. Yukawa and S. Amari, "ℓ_p-regularized least squares $(0 < p < 1)$ and critical path," *IEEE Trans. Information Theory*, vol. 62, no. 1, pp. 488–502, Jan. 2016.

[94] M. Yukawa and R. Ishii, "Online model selection and learning by multikernel adaptive filtering," in *Proc. EUSIPCO*, 2013.

[95] ——, "On adaptivity of online model selection method based on multikernel adaptive filtering," in *Proc. APSIPA Annual Summit and Conference*, 2013.

[96] O. Toda and M. Yukawa, "Online model-selection and learning for nonlinear estimation based on multikernel adaptive filtering," submitted for publication.

[97] M. Kasparick, R. L. G. Cavalcante, S. Valentin, S. Stanczak, and M. Yukawa, "Kernel-based adaptive online reconstruction of coverage maps with side information," *IEEE Trans. Vehicular Technology*, accepted.

[98] D. G. Luenberger, *Optimization by Vector Space Methods.* New York: Wiley, 1969.

[99] H. Q. Minh, "Some properties of Gaussian reproducing kernel Hilbert spaces and their implications for function approximation and learning theory," *Constr. Approx.*, vol. 32, no. 2, pp. 307–338, Oct. 2010.

[100] R. Vert and J. P. Vert, "Consistency and convergence rates of one-class SVMs and related algorithms," *J. Mach. Learn. Res.*, vol. 5, pp. 817–854, 2006.

[101] I. Steinwart, D. Hush, and C. Scovel, "An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4635–4643, Oct. 2006.

[102] A. Tanaka, H. Imai, M. Kudo, and M. Miyakoshi, "Theoretical analyses on a class of nested RKHS's," in *Proc. IEEE ICASSP*, 2011, pp. 2072–2075.

Department of Electronics and Electrical Engineering
Keio University
Kanagawa 223-8522
JAPAN
E-mail address: `yukawa@elec.keio.ac.jp`

慶應義塾大学・電子工学科　湯川　正裕