Lecture 4. Representer theorems on Hilbert spaces

In this lecture we present a simple, yet powerful, theoretical result concerning optimization problems on Hilbert spaces called Representer theorems. The results have wide applications in approx. theory, machine learning, statistics, & engineering.

Thm (Representer theorem)

Suppose $(H, \|\cdot\|, \langle \cdot, \cdot \rangle)$ is a Hilbert space & consider the optimization problem

$$(P) \quad \min_{h \in H} J(h) := L(\langle h, h_1 \rangle, \dots, \langle h, h_n \rangle) + R(\|h\|)$$

where $\{h_1, \dots, h_n\} \subset H$ are fixed & R is non-decreasing. Then if J admits minimizers then it has at least one minimizer h^* of the form

$$h^* = \sum_{j=1}^{n} \alpha_j h_j \qquad (RP)$$

A few important points:

• Note that we have minimal requirements on L beyond asking for J to admit at least one minimizer.

• The Thm simply states that J has at least one minimizer that belongs to $\operatorname{span}\{h_1, \dots, h_n\}$ which is n-dimensional! the h_j are sometimes called the representers of h^*

Proof of the representer theorem:

The proof is a consequence of a simple orthogonality condition. Let $\tilde{H} := \operatorname{span}\{h_1, \dots, h_n\}$ for any $h \in H$ we can write $h = h^{\parallel} + h^{\perp}$ where $h^{\parallel} \in \tilde{H}$ & $h^{\perp}$ belongs to the orthogonal complement of $\tilde{H}$. I.e., $\langle h^{\parallel}, h^{\perp} \rangle = 0$.

Observe that
$$\|h\|^2 = \langle h^{\parallel} + h^{\perp}, h^{\parallel} + h^{\perp} \rangle$$
$$= \|h^{\parallel}\|^2 + \|h^{\perp}\|^2 + \cancel{2\langle h^{\parallel}, h^{\perp} \rangle}_{0} \quad \} \; (*)$$

Suppose $\tilde{h}$ is a minimizer of J & write
$$\tilde{h} = \tilde{h}^{\parallel} + \tilde{h}^{\perp}$$

Now observe that

$$L(\langle \tilde{\hat{h}}, h_1 \rangle, \ldots, \langle \tilde{\hat{h}}, h_n \rangle)$$
$$= L(\langle \tilde{\hat{h}}^{\parallel}, h_1 \rangle + \langle \tilde{\hat{h}}^{\perp}, h_1 \rangle, \ldots, \langle \tilde{\hat{h}}^{\parallel}, h_n \rangle + \langle \tilde{\hat{h}}^{\perp}, h_n \rangle)$$
$$= L(\langle \tilde{\hat{h}}^{\parallel}, h_1 \rangle, \ldots, \langle \tilde{\hat{h}}^{\parallel}, h_n \rangle)$$

Therefore

$$J(\tilde{\hat{h}}) - J(\tilde{\hat{h}}^{\parallel}) = R(\|\tilde{\hat{h}}\|) - R(\|\tilde{\hat{h}}^{\parallel}\|) \geq 0$$

since $\|\tilde{\hat{h}}\| \geq \|\tilde{\hat{h}}^{\parallel}\|$ & $R: \mathbb{R} \to \mathbb{R}$ is non-decreasing!

Thus, $\tilde{\hat{h}}^{\parallel}$ is also a minimizer of J!

$\blacksquare$

<u>Corollary</u>: If R is strictly increasing then any minimizer of J is of the form (RP).

<u>Proof</u>: Just observe that in the last step of the proof we have $R(\|\tilde{\hat{h}}\|) > R(\|\tilde{\hat{h}}^{\parallel}\|)$, if $\tilde{\hat{h}}^{\perp} \neq 0$ leading to a contradiction.

$\blacksquare$

Representer thms are powerful tools because Problems of the form (P) are commonly encountered in practice.

eg: (Regularized least squares on $\mathbb{R}^n$)

$\underline{h}_1, \ldots, \underline{h}_n \in \mathbb{R}^d$, $d \gg n$

opt for $h \in \mathbb{R}^d$

$$J(\underline{h}) = \sum_{j=1}^{n} |\underline{h}^T \underline{h}_j - y_j|^2 + R(\|\underline{h}\|_2)$$

$$= \|A\underline{h} - \underline{y}\|_2^2 + R(\|\underline{h}\|_2)$$

By rep. thm we have that $\hat{\underline{h}} = \sum_{j=1} \alpha_j \underline{h}_j$, sub in J to get,

opt for $\underline{\alpha} \in \mathbb{R}^n$

$$J(\underline{\alpha}) = \sum_{j=1}^{n} |\alpha_j \cdot \|\underline{h}_j\|_2^2 - y_j|^2 + R(\|\sum_{j=1}^{n} \alpha_j \underline{h}_j\|_2^2)$$

eg: (Optimal Recovery in $L^2([-\pi,\pi])$)

Let $h_0 = \frac{1}{\sqrt{2\pi}}$, $h_j(x) = \frac{1}{\sqrt{\pi}} \cos(j\pi x)$ for $j = 1, \ldots, n$.

consider the optimization problem,

opt. over L^2!

$$J(h) = \sum_{j=0}^{n} \left| \int_{-\pi}^{\pi} h(t) h_j(t) \, dt - y_j \right| + \|h\|^2_{L^2(-\pi,\pi)}$$

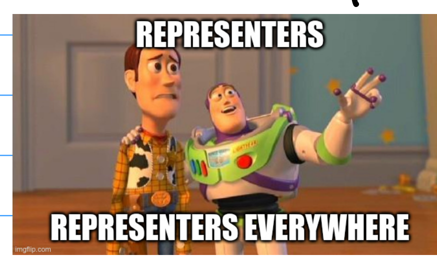
(4) Apply rep. thm. $\hat{h} = \sum_{j=0}^{n} \alpha_j h_j$

in fact, here the h_j are an orthonormal set & so,

n-dim opt. problem!

$$J(\underline{\alpha}) = \sum_{j=0}^{n} | \alpha_j - y_j | + \|\underline{\alpha}\|_2^2$$

While the representer theorem is quite powerful it may seem limited at first, since we require the "measurements" to be of the form $\langle h, h_j \rangle$. However, the result can be generalized to all bounded linear functionals on H thanks to Riesz's representation theorem. (This is a diff. representation thm).



<u>Thm</u> (Riesz's representation theorem)
Every bounded linear functional ϕ on a Hilbert space H (ie, $\phi \in H^*$) can be represented in terms of the H-inner product

$$\phi(h) = \langle \hat{\phi}, h \rangle$$

5 where $\hat{\phi} \in H$ depends on ϕ. it is uniquely

determined by ϕ & satisfies $\|\hat{\phi}\| = \|\phi\|_*$

(we will discuss the proof next lecture).

Since Riesz's rep. thm. allows us to identify bdd lin. func. with inner product with elements of the Hilbert space. We can now generalize our Rep. Thm. to the following form.

<u>Thm</u> Let $(H, \|\cdot\|, \langle\cdot,\cdot\rangle)$ be a Hibert space & consider the opt problem

$$J(h) = L(\phi_1(h), \ldots, \phi_n(h)) + R(\|h\|)$$

where $\phi_j \in H^*$ are fixed & $R : \mathbb{R} \to \mathbb{R}$ is non-decreasing. Then if J admits minimizers then it has at least one minizer $\hat{h}$ of the form

$$\hat{h} = \sum_{j=1}^{n} \alpha_j \hat{\phi}_j$$

where the $\hat{\phi}_j$ are the Riesz representers of the ϕ_j.

6 <u>Proof:</u> Apply Riesz's rep to write $\phi_j(h) = \langle h, \hat{\phi} \rangle$ & apply the original rep thm.

This form of the Rep. thm. is very useful in our study of RKHS's where the $\hat{\phi}$ often have simple & convenient forms in terms of a Kernel function.