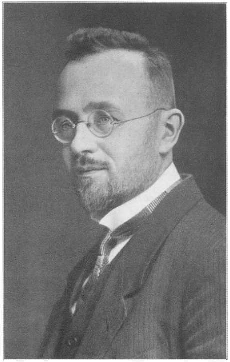
Regularized Radon-Nikodym differentiation and some of its applications

Sergei Pereverzyev

Joint research with Nguyen Duc Hoan (RICAM) and Werner Zellinger (RICAM) performed within FFG COMET project S3AI.

Workshop 5
RICAM Special Semester 2022.

Otton Marcin Nikodym (1887 - 1974)



Johann Radon (1887 - 1956)

A motivation for the study

- Supervised learning paradigm:
  a) We observe input-output relationship $x \to y, x \in \mathbf{X}, y \in \mathbf{Y}$, governed by a probabilistic law with probability measure $p(x, y)$;
  b) a learning algorithm predicts y for previously unseen x as a value $f_\theta(x)$ of some model function $f_\theta : \mathbf{X} \to \mathbf{Y}$ parametrized by a parameter (vector) θ;
  c) the expected error (risk) of the prediction $y = f_\theta(x)$ is defined as

$$\mathbb{E}_p(f_\theta) = \int_{\mathbf{X}\times\mathbf{Y}} \mathbf{e}(f_\theta(x), y) dp(x, y),$$

  where $\mathbf{e}(f_\theta(x), y)$ is some error measure, such as, for example, $\mathbf{e}(f(x), y) = \|f(x) - y\|_{\mathbf{Y}}^2$;

A motivation for the study(continuation)

- Supervised learning paradigm:
  d) We are provided with a training set of n previously observed input-output pairs $z = \{(x_i, y_i)\}_{i=1}^n$, which are assumed to be i.d.d drawn from $p(x, y)$;
  e) the empirical risk minimization principle states that the learning algorithm should choose θ that minimizes the empirical risk

$$\mathbb{E}_{z,p}(f_\theta) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}(f_\theta(x_i), y_i),$$

  or its regularized/penalized version.

Domain adaptation problem

- Can we use training set $z = \{(x_i, y_i)\}_{i=1}^n$ drawn from (the source measure) $p(x, y)$ for predicting input-output relationship $x \to y$ governed by another (target) probability measure $q(x, y)$?
- The covariate shift assumption (H. Shimodaira, 2000): both source and target measures share the same conditional measure, say $\rho(y|x)$, while their marginal measures, $\rho_S(x)$ and $\rho_T(x)$, are different,i.e.

$$p(x, y) = \rho(y|x)\rho_S(x), \quad q(x, y) = \rho(y|x)\rho_T(x).$$

Domain adaptation problem (continuation)

Example: x contains the prescribed health-related measurements observed in two different areas $(\rho_S(x) \neq \rho_T(x))$ in relation to patients at risk, y, of the same pathology.

- Key assumption (J. Huang et. al., 2006): Existence of the Radon-Nikodym derivative $\beta : \mathbf{X} \to \mathbb{R}_+$, $\beta(x) = \frac{d\rho_T(x)}{d\rho_S(x)}$ such that

$$d\rho_T(x) = \beta(x) d\rho_S(x).$$

Sample Reweighting

- Covariate shift assumption + Key assumption allow for the following relation:

$$\mathbb{E}_q(f_\theta) = \int_{\mathbf{X}\times\mathbf{Y}} \mathbf{e}(f_\theta(x), y) dq(x, y) = \int_{\mathbf{X}\times\mathbf{Y}} \mathbf{e}(f_\theta(x), y)\beta(x) dp(x, y).$$

- Knowing $\beta(x_i), i = 1, 2, \ldots, n$, we can approach the domain adaptation problem by performing the empirical risk minimization with training data $z = \{(x_i, y_i)\}_{i=1}^n$ drawn from the source measure $p(x, y)$:

$$\mathbb{E}_{z,q}(f_\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}(f_\theta(x_i), y_i)\beta(x_i) \to \min .$$

Sample Reweighting (continuation)

- This allows for unsupervised domain adaptation, because no examples of input-output pairs drawn from the target measure $q(x, y)$ are required.

Radon-Nikodym (R-N) differentiation in RKHS

- Since we are interested in knowing the point values of $\beta = \frac{d\rho_T}{d\rho_S}$ at $x_i \in \mathbf{X}$, it is natural to assume that β belongs to a space of functions $\mathcal{H}$, where pointwise evaluation is well-defined as a continuous linear functional. Then $\mathcal{H} = \mathcal{H}_K$ is Reproducing Kernel Hilbert space (RKHS) generated by a symmetric and positive definite kernel $K : \mathbf{X} \times \mathbf{X} \to \mathbb{R}_+$.
- Recall that RKHS $\mathcal{H}_K$ is the completion of the space of the linear combinations of kernel sections $K_x(\cdot) = K(\cdot, x)$, $x \in \mathbf{X}$, with respect to the inner product $\langle \cdot, \cdot \rangle_K$ for which $\langle K_x, K_t \rangle_K := K(x, t)$. Moreover, for any $f \in \mathcal{H}_K$ and $x \in \mathbf{X}$ we have

$$f(x) = \langle K_x, f \rangle_K .$$

Basic equation for R-N differentiation in RKHS

- By the definition of $\beta = \frac{d\rho_T}{d\rho_S}$ it follows that for any $f \in \mathcal{H}_K$

$$\int_{\mathbf{X}} f(t)d\rho_T(t) = \int_{\mathbf{X}} f(t)\beta(t)d\rho_S(t),$$

and, in particular, for $f(\cdot) = K_t(\cdot) = K(\cdot, t)$ we have

$$\int_{\mathbf{X}} K(\cdot, t)d\rho_T(t) = \int_{\mathbf{X}} K(\cdot, t)\beta(t)d\rho_S(t).$$

- If we consider the canonical embedding operators $J_T : \mathcal{H}_K \hookrightarrow L_{2,\rho_T}$, $J_S : \mathcal{H}_K \hookrightarrow L_{2,\rho_S}$, then it can be shown that

$$J_T^* f(\cdot) = \int_{\mathbf{X}} K(\cdot, t)f(t)d\rho_T(x), \quad J_S^* f(\cdot) = \int_{\mathbf{X}} K(\cdot, t)f(t)d\rho_S(x)$$

Basic equation (continuation)

- Assume that the constant function $\mathbf{1}(\cdot) \equiv 1$ belongs to $\mathcal{H}_K$. Then $\beta = \frac{d\rho_T}{d\rho_S} \in \mathcal{H}_K$ should solve the first kind Fredholm integral equation with compact and self-adjoint integral operator

$$J_S^* J_S \beta = J_T^* J_T \mathbf{1},$$

which is known to be an ill-posed problem and can be treated only with regularization technique.

- According to P. Mathe and B. Hofmann (2008) the smoothness properties of $\beta = \frac{d\rho_T}{d\rho_S}$ can always be described in terms of general source conditions. Namely, for any $\epsilon > 0$ there is a continuous strictly increasing function $\varphi : [0, c] \to \mathbb{R}_+$, $c > \|J_S^* J_S\|$, such that $\varphi(0) = 0$ and

$$\beta = \varphi(J_S^* J_S)\vartheta, \qquad \vartheta \in \mathcal{H}_K, \|\vartheta\|_{\mathcal{H}_K} < (1+\epsilon)\|\beta\|_{\mathcal{H}_K}.$$

Basic equation (continuation)

Example: If $\mathcal{H}_K$ is a Sobolev space $W_2^s \hookrightarrow C(X)$, and $\mathbf{X}$ is a closed smooth manifold, then $\beta = \varphi(J_S^* J_S)\vartheta \in W_2^{s,\varphi}(\mathbf{X})$, where $W_2^{s,\varphi}(\mathbf{X})$ is the so-called refined Sobolev scale (A. Mikhailets, A. Murach,2012), which is much finer than the standard Sobolev scale.

Monte - Carlo quadrature method

- Note that the equation $J_S^* J_S \beta = J_T^* J_T \mathbf{1}$ is inaccesible, because we do not know marginal measures ρ_T, ρ_S. In practice, we are provided with samples $\{x_i\}_{i=1}^n, \{x_j'\}_{j=1}^m$ of unlabeled examples of inputs without knowing the corresponding outputs.
- The samples $\{x_i\}_{i=1}^n, \{x_j'\}_{j=1}^m$ are supposed to be i.i.d drawn corresponding from ρ_S and ρ_T, and we consider sampling operators $S_{n,S} : \mathcal{H}_K \to \mathbb{R}^n$, $S_{m,T} : \mathcal{H}_K \to \mathbb{R}^m$, associated to them, i.e.,

$$S_{n,S} f = (f(x_1), f(x_2), \dots, f(x_n)),$$
$$S_{m,T} f = (f(x_1'), f(x_2'), \dots, f(x_m')).$$

Monte - Carlo quadrature method (continuation)

- By using Monte-Carlo quadrature formulas we can approximate left and right sides of the basic equation as follows:

$$J_S^* J_S \beta(\cdot) = \int_{\mathbf{X}} K(\cdot, t)\beta(t) d\rho_S(t)$$

$$\approx \frac{1}{n}\sum_{i=1}^{n} K(\cdot, x_i)\beta(x_i) = S_{n,S}^* S_{n,S}\beta(\cdot),$$

$$J_T^* J_T \mathbf{1} = \int_{\mathbf{X}} K(\cdot, t) d\rho_T(t) = \frac{1}{m}\sum_{j=1}^{m} K(\cdot, x_j') = S_{m,T}^* S_{m,T}\mathbf{1},$$

and this gives us a discretized version of R-N differentiation problem:

$$S_{n,S}^* S_{n,S}\beta = S_{m,T}^* S_{m,T}\mathbf{1}$$

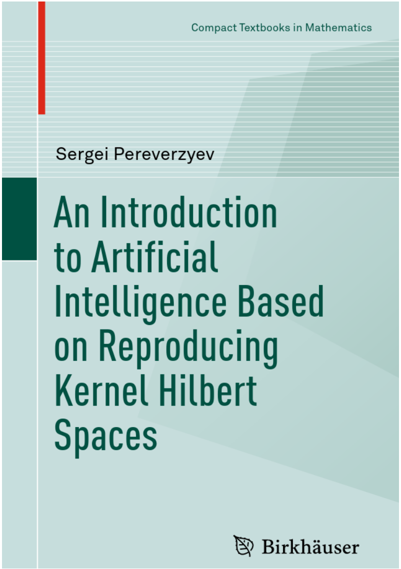
Application of the regularization theory

- Arguments based on concentration of measures allow for the estimation of the amount of discretization noise. Namely, with probability at least $1-\delta$ we have

$$\left\|J_S^* J_S - S_{n,S}^* S_{n,S}\right\|_{\mathcal{H}_K \to \mathcal{H}_K} \leq c n^{-\frac{1}{2}} \log^{\frac{1}{2}}\left(\frac{1}{\delta}\right),$$

$$\left\|J_T^* J_T \mathbf{1} - S_{n,T}^* S_{n,T} \mathbf{1}\right\|_{\mathcal{H}_K \to \mathcal{H}_K} \leq c m^{-\frac{1}{2}} \log^{\frac{1}{2}}\left(\frac{1}{\delta}\right),$$

where here and below c denotes a generic positive coefficient that does not depend on the quantities of interest. The above bounds open the way for straightforward use of the arguments developed in the regularization theory.


Compact Textbooks in Mathematics
Sergei Pereverzyev
An Introduction to Artificial Intelligence Based on Reproducing Kernel Hilbert Spaces
Birkhäuser

Application of the regularization theory (cont.)

- Recall that the procedure , in which approxiamte solutions f_λ of equation $Tf = u, \quad T = T^* \geq 0,$ are constructrd as $f_\lambda = g_\lambda(T)u$ is called the regularization indexed by a family of functions $\{g_\lambda(t)\}$ if $\exists \gamma_0, \gamma_{-1} > 0$:

$$\sup_{0 \leq t \leq \|T\|} |1 - g_\lambda(t)t| \leq \gamma_0, \qquad \sup_{0 \leq t \leq \|T\|} |g_\lambda(t)| \leq \frac{\gamma_{-1}}{\lambda}.$$

- Qualification of the regularization indexed by $\{g_\lambda\}$ is the maximal p for which $\exists \gamma_p > 0$:

$$\sup_{0 \leq t \leq \|T\|} t^p |1 - g_\lambda(t)t| \leq \gamma_p \lambda^p.$$

Application of the regularization theory (cont.)

- We say that the qualification p covers the index function $\varphi : [0, \|T\|] \to \mathbb{R}_+$, $\varphi(0) = 0$, if the function $t \to t^p/\varphi(t)$ is non-decreasing on $(0, ||T||]$.

Example: The so-called Lavrentiev regularization is indexed by $\{g_\lambda(t) = (\lambda + t)^{-1}\}$ and has qualification $p = 1$; k-times iterated Lavrentiev regularization is indexed by $\{g_\lambda(t) = \frac{(1-(\lambda/(\lambda+t)))^k}{t}\}$ and has qualification $p = k$.

Theorem 1

Let $\beta(x) = \frac{d\rho_T(x)}{d\rho_S(x)} = \varphi(J_S^* J_S)\vartheta \in \mathcal{H}_K$. *Consider the approximant*

$$\beta_{m,n}^{\lambda} = g_\lambda(S_{n,S}^* S_{n,S}) S_{m,T}^* S_{m,T} \mathbf{1},$$

where $\{g_\lambda\}$ *has the qualification* p *that covers* $\varphi(t)$. *Consider also* $\theta_\varphi(t) = \varphi(t)t$ *and* $\lambda_{m,n} = \theta_\varphi^{-1}\left(m^{-\frac{1}{2}} + n^{-\frac{1}{2}}\right)$. *Then for sufficiently large* m *and* n *with probability at least* $1-\delta$ *it holds*

$$\left\|\beta - \beta_{m,n}^{\lambda_{m,n}}\right\|_{\mathcal{H}_K} \le c\varphi(\lambda_{m,n}) \log\left(\frac{1}{\delta}\right).$$

If, in addition, the qualification p *covers* $\varphi(t)\sqrt{t}$, *then*

$$\left\|\beta - \beta_{m,n}^{\lambda_{m,n}}\right\|_{L_{2,\rho_S}} \le c\varphi(\lambda_{m,n})\sqrt{\lambda_{m,n}} \log\left(\frac{1}{\delta}\right).$$

Example and comparisions

- T. Kanamori et al. (2012) have proposed the so-called kernelized unconstrained least squares importance fitting (KuLSIF) for approximating $\beta(x) = \frac{d\rho_T(x)}{d\rho_S(x)} \in \mathcal{H}_K$ from samples $\{x_i\}_{i=1}^n, \{x_j'\}_{j=1}^m$ i.i.d drawn from ρ_S and ρ_T. In our terms, KuLSIF is nothing but Lavrentiev regularization indexed by $\{g_\lambda(t) = (\lambda + t)^{-1}\}$.
- In spite of the fact there always is some φ such that $\beta = \varphi(J_S^* J_S)\vartheta$, T. Kanamori et all. (2012) did not take into account any smoothness of $\beta \in \mathcal{H}_K$. Instead, the bound $\left\|\beta - \beta_{m,n}^\lambda\right\|_{L_{2,\rho_S}}$ was established in terms of the order γ of the so-called bracketing entropy of $\mathcal{H}_K$. In such terms the best proven bound was order $O\left((m \wedge n)^{-\frac{1}{2+\gamma}}\right), 0 < \gamma < 2$.

- Observe that, for example, for $\varphi(t) = t^q, q > \frac{1}{\gamma} - \frac{1}{2}$, the above order is worse than the one given by Theorem 1, but the idea to take into account an entropy of underlying $\mathcal{H}_K$ goes beyond the standard regularization theory, and will be discussed below.
- KuLSIF has also been analyzed by I. Schuster et al. (2020). In our terms, the corresponding result can be written as follows:

$$\left\|\beta - \beta_{m,n}^{\lambda}\right\|_{\mathcal{H}_K} \leq c\left(\varphi(\lambda) + \frac{n^{-a}}{\lambda^2} + \frac{m^{-b}}{\lambda}\right)\log\frac{1}{\delta},$$

where $a, b < \frac{1}{2}$, while a straightforward application of the regularization theory can give a better bound

$$\left\|\beta - \beta_{m,n}^{\lambda}\right\|_{\mathcal{H}_K} \leq c\left(\varphi(\lambda) + \frac{m^{-\frac{1}{2}} + n^{-\frac{1}{2}}}{\lambda}\right)\log\frac{1}{\delta}.$$

(Regularized) Christoffel functions

- Theorem 1 is currently a state of the art in theoretical approximation results for R-N numerical differentiation. It is valid for any $\mathcal{H}_K$ and ρ_T, ρ_S such that $\frac{d\rho_T}{d\rho_S} \in \mathcal{H}_K$, but it does not account for a specific interplay between RKHS and the considered measures.
- It is know that, in case of $\mathcal{H}_{K_l}$ the finite dimensional RKHS of polynomials of degree at most l, the classical Christoffel function

$$C_{K_l,\rho_S}(x) = \min\left\{ \int_{\mathbf{X}} f^2(t) d\rho_S(t) : f \in \mathcal{H}_{K_l}, f(x) = 1 \right\}$$

is a proper term for describing the above interplay.

R-N meet Christoffel, at least conceptually

Otton Marcin Nikodym (1887 - 1974)

Erwin Bruno Christoffel (1829 - 1900)

Johann Radon (1887 - 1956)

R-N meet Christoffel, at least conceptually

- If x belongs to the interior of the common support of ρ_T, ρ_S then (A. Kroó, D.S. Lubinsky, 2012)

$$\lim_{l\to\infty} \frac{C_{K_l,\rho_T}(x)}{C_{K_l,\rho_S}(x)} = \frac{d\rho_T}{d\rho_S}(x).$$

(Regularized) Christoffel functions

- The regularized Christoffel function (E. Puwels et al., 2018)

$$C^{\lambda}_{K,\rho_S}(x) = \inf\left\{ \int_{\mathbf{X}} f^2(t) d\rho_S(t) + \lambda ||f||^2_K : f(x) = 1 \right\}$$

  is a direct extension of the classical Christoffel function to the case of infinite dimentional RKHS.
- Asymptotic behaviour of $C_{K,\rho_S}(x)$ as $\lambda \to 0$ has been analysed (E. Panwels et al., 2018) for translation invariant kernels $K(x,t) = K(x-t)$. Here we describe such behavior in terms of general source conditions on kernel sections $K_t(\cdot) = K(\cdot,t) \in \mathcal{H}_K$.

Christoffel functions and source conditions

- As we know (P. Mathe, B. Hofmann, 2008) $\forall t \in \mathbf{X}, \exists \psi_t : \psi_t(0) = 0, \psi^{\nearrow}, K_t = \psi_t(J_S^* J_S)\vartheta_t, \vartheta_t \in \mathcal{H}_K$. We follow E. De Vito et al., 2014, and S. Lu et al. 2018, and consider a majorant of all ψ_t.
  Assumption 1: $\exists \psi : \psi(0) = 0,\ \psi^{\nearrow}, \forall t \in \mathbf{X}$

$$K(\cdot, t) = \psi(J_S^* J_S)\vartheta_t,$$

  $||\vartheta_t||_K \leq (1+\epsilon)||K_t||_K \leq (1+\epsilon)\varkappa,\ \varkappa = \max\{K(t,t), t \in \mathbf{X}\}$.
  Moreover, $\psi^2(t)$ is covered by the qualification $p = 1$.

Christoffel functions and source conditions

- From (E. Pauwels et al. 2018) we know that

$$C^{\lambda}_{K,\rho_S}(x) = \left(\left\langle K_x, (\lambda I + J_S^* J_S)^{-1} K_x \right\rangle_K \right)^{-1}.$$

On the other hand, the functions

$$\mathcal{N}_x(\lambda) = \left\langle K_x, (\lambda I + J_S^* J_S)^{-1} K_x \right\rangle_K$$

and $\mathcal{N}_\infty(\lambda) = \sup\{\mathcal{N}_x(\lambda), x \in \mathbf{X}\}$ play an important role in the analysis of regularized learning in RKHS (A. Rudi et al. 2015), (S. Lu et al., 2018). The above functions can be called reciprocal of the regularized Christoffel functions, and Assumption 1 allows us to estimate them as follows:

$$\mathcal{N}_x(\lambda) \leq \mathcal{N}_\infty(\lambda) \leq c\frac{\psi^2(\lambda)}{\lambda}.$$

More tight bounds for R-N numerical differentiation

Theorem 2

If, under the assumptions of Theorem 1, the qualification of the regularization indexed $\{g_\lambda\}$ *covers the function* $\varphi(t)t$ *then with probability at least* $1-\delta$ *it holds*

$$\left\|\beta - \beta^{\lambda}_{m,n}\right\|_{\mathcal{H}_K} \leq c\left(\varphi(\lambda) + \sqrt{\frac{N_\infty(\lambda)}{\lambda(m+n)}}\right)\log^2\frac{1}{\delta},$$

where λ *is supposed to be such that* $\frac{N_\infty(\lambda)}{\lambda(m+n)} \leq 1$.

More tight bounds for R-N numerical differentiation

Corollary 2.1

Let Assumption 1 and the assumptions of Theorem 2 be satisfied. Consider $\theta_{\varphi,\psi}(t) = \varphi(t)t/\psi(t)$ *and* $\bar{\lambda}_{m,n} = \theta^{-1}_{\varphi,\psi}(m^{-\frac{1}{2}} + n^{-\frac{1}{2}})$. *Then with probability at least* $1-\delta$ *it holds*

$$\left\|\beta - \beta^{\lambda}_{m,n}\right\|_{\mathcal{H}_K} \leq c\varphi(\bar{\lambda}_{m,n}) \log^2 \frac{1}{\delta}.$$

Note that the error bound given by Corollary 2.1 is of higher order of smallness than the one given by Theorem 1.

Pointwise approximation of R-N derivatives

- Recall that performing the empirical risk minimization in the context of the domain adaptation, we need to know not the whole R- N derivative $\beta = \frac{d\rho_T}{d\rho_S}$, but only its values at the inputs $x_1, x_2, \ldots, x_n$.
- In view of Amssuption 1

$$\begin{aligned} \left|\beta(x_i) - \beta^{\lambda}_{m,n}(x_i)\right| &= \left|\left\langle K_{x_i}, \beta - \beta^{\lambda}_{m,n}\right\rangle_K\right| \\ &= \left|\left\langle \vartheta_{x_i}, \psi(J_S^* J_S)(\beta - \beta^{\lambda}_{m,n})\right\rangle\right| \\ &\leq 2\varkappa \left\|\psi(J_S^* J_S)(\beta - \beta^{\lambda}_{m,n})\right\|_{\mathcal{H}_K}, \end{aligned}$$

i.e., we need to estimate the error of R-N numerical differentiation in a weighted norm associated with $\psi(J_S^* J_S)$.

Pointwise approximation of R-N derivatives

Theorem 3

If, under the assumptions of Corollary 2.1, the qualification of the regularization indexed by $\{g_\lambda\}$ *covers the function* $\varphi(t)t^{\frac{3}{2}}$ *then for any* $x_i \in \mathbf{X}$ *with probability at least* $1-\delta$ *it holds*

$$\left|\beta(x_i) - \beta^{\lambda}_{m,n}(x_i)\right| \leq c\psi(\lambda)\left(\varphi(\lambda) + \frac{\psi(\lambda)}{\lambda\sqrt{m+n}}\right)\log^2\frac{1}{\delta},$$

and for $\lambda = \bar{\lambda}_{m,n} = \theta^{-1}_{\varphi,\psi}(m^{-\frac{1}{2}} + n^{-\frac{1}{2}})$

$$\left|\beta(x_i) - \beta^{\lambda}_{m,n}(x_i)\right| \leq c\psi(\bar{\lambda}_{m,n})\varphi(\bar{\lambda}_{m,n})\log^2\frac{1}{\delta},$$

i.e., $\beta(x_i)$ *can be estimated much more accurately than* β *in* $\mathcal{H}_K$.

Numerical illustrations

- Theorem 3 hints that the same value of the regularization parameter λ can be used for approximating all $\beta(x_i)$, $i = 1, 2, \ldots, n$. Therefore, in practice one can use the so-called quasi-optimality criterion to choose $\bar{\lambda} \in \{\lambda_k = \lambda_0 q^k, k = 1, 2, \ldots, l\}, q > 1$ such that for $\bar{\lambda} = \lambda_{k_0}$

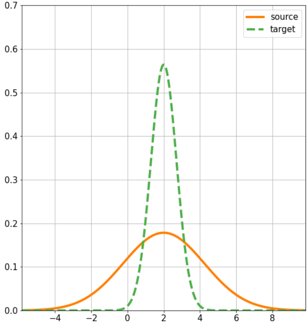
$$\left\|\beta_{m,n}^{\lambda_{k_0}} - \beta_{m,n}^{\lambda_{k_0-1}}\right\|_{\mathbb{R}^n} = \min\left\{\left\|\beta_{m,n}^{\lambda_k} - \beta_{m,n}^{\lambda_{k-1}}\right\|_{\mathbb{R}^n}, k = 1, 2, \ldots, l\right\},$$

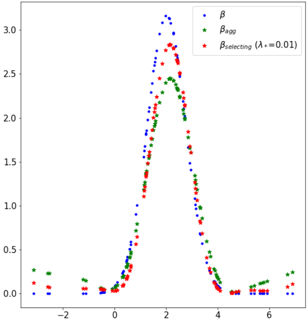
  where $\beta_{m,n}^{\lambda} = (\beta_{m,n}^{\lambda}(x_1), \beta_{m,n}^{\lambda}(x_2), \ldots, \beta_{m,n}^{\lambda}(x_n)) \in \mathbb{R}^n$.
- Quasi-optimality criterion can also be used to choose λ for approximating β in $\mathcal{H}_K$, i.e. $\bar{\lambda} = \lambda_{k_1}$:

$$\left\|\beta_{m,n}^{\lambda_{k_1}} - \beta_{m,n}^{\lambda_{k_1-1}}\right\|_{\mathbb{R}^n} = \min\left\{\left\|\beta_{m,n}^{\lambda_k} - \beta_{m,n}^{\lambda_{k-1}}\right\|_{\mathcal{H}_K}, k = 1, 2, \ldots, l\right\}.$$

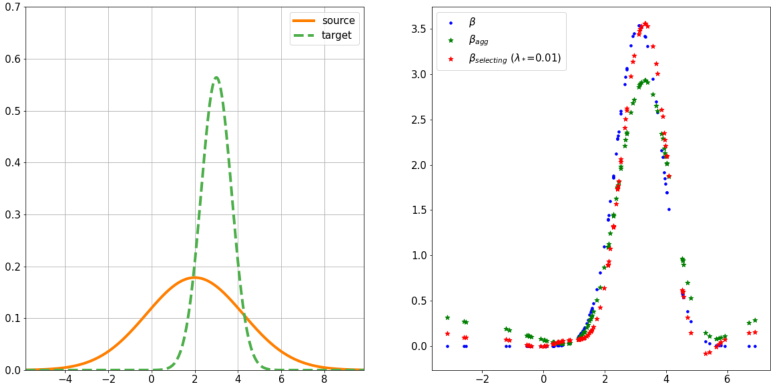
Numerical illustrations (continuation)

Numerical simulation with the above criteria support the theoretical conclusion coming from Theorem 2 and 3.

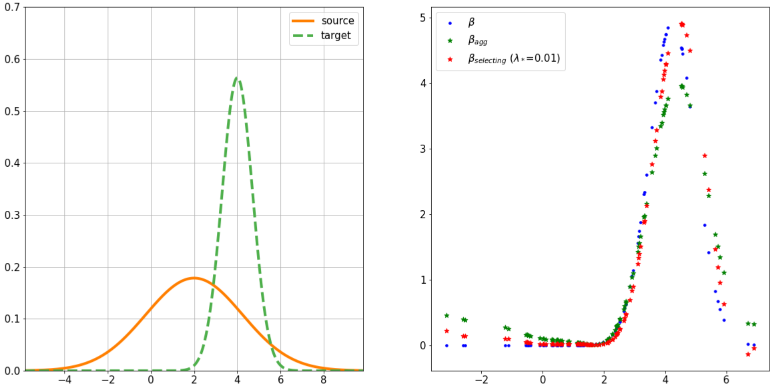




Numerical illustrations (continuation)

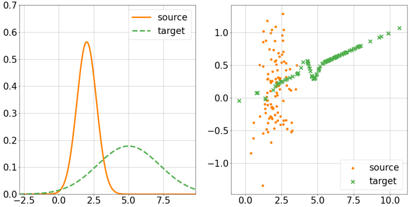


Numerical illustrations (continuation)



R-N in Regularized Empirical risk minimization

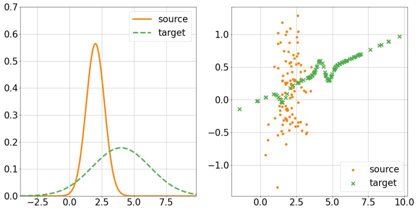
Source distribution: $N(2, 0.5)$, Noise: $N(0, 0.25)$,
Target distribution: $N(5, 5)$



Random seed	MSE $(\beta_{m,n}^{\lambda})$	MSE $(\beta = 1)$
1	**0.0146**	0.0241
64	**0.0052**	0.0072
99882	**0.0591**	0.1631

R-N in Regularized Empirical risk minimization

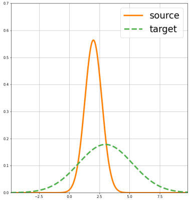
Source distribution: $N(2, 0.5)$, Noise: $N(0, 0.25)$,
Target distribution: $N(4, 5)$

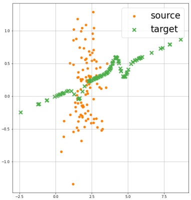


Random seed	MSE $(\beta^{\lambda}_{m,n})$	MSE $(\beta = 1)$
1	**0.0288**	0.0375
64	**0.0059**	0.0067
99882	**0.1004**	0.1402

R-N in Regularized Empirical risk minimization

Source distribution: $N(2, 0.5)$, Noise: $N(0, 0.25)$,
Target distribution: $N(3, 5)$





Random seed	MSE $(\beta_{m,n}^{\lambda})$	MSE $(\beta = 1)$
1	0.0378	**0.0291**
64	**0.0069**	0.0071
99882	**0.0828**	0.0855