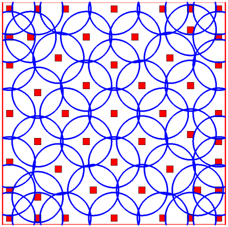
OLSE and BLUE for the location scale model and energy minimization

Anatoly Zhigljavsky, Cardiff (joint work with Luc Pronzato)



Workshop on kernel approximations and space-filling,
Cardiff, July 1, 2022

Energy

Let $\mathcal{X}$ be a (compact) set and $K(x, x')$ be a continuous kernel.
For any signed measure μ on $\mathcal{X}$, its energy is

$$\mathcal{E}(\mu) = \int_{\mathcal{X}} \int_{\mathcal{X}} K(x, x') \mu(dx) \mu(dx') .$$

If the kernel K is IPD (integrally positive definite), then $\mathcal{E}(\mu) \geq 0$ for any μ and $\mathcal{E}(\mu)$ is a convex functional of μ (conditions on K can be relaxed).

Minimum-energy signed measure of mass 1: $\mu^* = \arg\min_{\mu \in \mathcal{M}} \mathcal{E}(\mu)$, where $\mathcal{M}$ is the set of signed measures μ with $\mu(\mathcal{X}) = 1$.

Minimum-energy probability measure: $\mu^+ = \arg\min_{\mu \in \mathcal{M}^+} \mathcal{E}(\mu)$.

μ^+ always exists but μ^* exists very rarely.

Maximum mean discrepancy: $\mathrm{MMD}(\mu, \nu) = \sqrt{\mathcal{E}(\mu - \nu)}$

Minimum-energy measures: optimality conditions

Theorem

(i) μ is a minimum-energy probability measure iff $P_\mu(x) \geq \mathcal{E}(\mu)$ for all $x \in \mathcal{X}$ and $P_\mu(x) = \mathcal{E}(\mu)$ on the support of μ.
(ii) μ is a minimum-energy signed measure of total mass one iff $P_\mu(x) = \mathcal{E}(\mu)$ for all $x \in \mathcal{X}$.

Here $P_\mu(x) = \int K(x, x')\mu(dx')$ is the potential (kernel imbedding) of μ.

Corollary

1 is a potential of some positive measure iff $\mu^+ = \mu^*$.
1 is a potential of a finite signed measure iff μ^* exists.

Questions:

When does 1 belong to the space of potentials?

If it does, when the corresponding measure is positive?

When minimum-energy measure is a probability measure?

Corollary

Let μ^+ be a minimum-energy probability measure (always exists!). If μ^+ is supported on the whole set $\mathcal{X}$, then μ^+ is also a minimum-energy signed measure of total mass one.

Theorem (PZ, 2020)

Let K be ISPD and translation invariant, with $K(x, x') = \psi(x - x')$ and ψ continuous, twice differentiable except at the origin, with Laplacian $\Delta\psi(x) = \sum_{i=1}^{d} \partial^2\psi(x)/\partial x_i^2 \geq 0$, $\forall x \neq 0$. Then there exists a unique mass 1 minimum-energy signed measure, which is a probability measure.

Idea of proof: potential P_{μ^*} is subharmonic outside the support of μ^*.
Particular case: $d = 1$, $\psi(x)$ is convex for $x > 0$ (Hàjek, 1956).

It is easy to construct translation invariant kernels for $d = 1$ such that $\psi(x)$ is not convex for $x > 0$ but mass 1 minimum-energy signed measure μ^* is a probability measure (help of T. Karvonen is appreciated).

OLSE, n-point design

Design: $X_n = \{x_1, \ldots, x_n\} \in \mathcal{X}$

Model: $y(x_j) = \theta + \varepsilon(x_j), \quad \mathbb{E}\varepsilon(x) = 0$, $\mathbb{E}\varepsilon(x)\varepsilon(x') = K(x, x')$,

K is a (conditionally) positive definite kernel

OLSE of θ:

$$\hat{\theta}_{OLSE,n} = \int y(x)\mu_n(dx) = \bar{y} = \frac{1}{n} 1_n^\top Y,$$

where $Y = (y(x_1), \ldots, y(x_n))^\top$, $1_n = (1, \ldots, 1)^\top$ and μ_n is the empirical probability measure assigning weights $1/n$ to points $x_j \in X_n$.

Variance of OLSE, n-point design

$$\begin{aligned} \mathrm{var}(\hat{\theta}_{OLSE,n}) &= \frac{1}{n^2} 1_n^\top K_n 1_n = \frac{1}{n^2} \sum_{i,j=1}^{n} K(x_i, x_j) \\ &= \int \int K(x, x') \mu_n(dx) \mu_n(dx'), \end{aligned}$$

where $K_n = (K(x_i, x_j))_{i,j=1}^n$ is the kernel matrix

That is,

$$\mathrm{var}(\hat{\theta}_{OLSE,n}) = \int \int K(x, x') \mu_n(dx) \mu_n(dx') = \mathcal{E}(\mu_n),$$

which is a discrete energy.

OLSE: approximate design; optimal designs

Approximate design: any probability measure μ on $\mathcal{X}$

$$\hat{\theta}_{OLSE} = \int y(x)\mu(dx)$$

$$\mathrm{var}(\hat{\theta}_{OLSE}) = \int\int K(x,x')\mu(dx)\mu(dx') = \mathcal{E}(\mu)$$

Minimum-energy probability measure $\mu^+ = \arg\min_{\mu\in\mathcal{M}^+} \mathcal{E}(\mu)$ is the optimal approximate design for OLSE (easy to construct numerically).

Optimal n-point design for OLSE is the minimum-energy n-point probability measure (hard to construct numerically).

BLUE of θ, n-point design

Design: $X_n = \{x_1, \ldots, x_n\} \in \mathcal{X}$, x_j are pair-wise different points.

$$\hat{\theta}_{BLUE,n} = w_n^* Y; \;\; w_n^* = 1_n^\top K_n^{-1} / 1_n^\top K_n^{-1} 1_n; \;\; \mathrm{var}(\hat{\theta}_{BLUE,n}) = 1/1_n^\top K_n^{-1} 1_n$$

w_n^* gives the weights of μ_n^*, the optimal signed measure minimizing the discrete energy

$$\mathcal{E}(\nu_n) = \int \int K(x, x') \nu_n(dx) \nu_n(dx'),$$

where ν_n are discrete signed measures supported on X_n with $\nu_n(X_n) = 1$.

$$\hat{\theta}_{BLUE,n} = \int y(x) \mu_n^*(dx), \;\; \mathrm{var}(\hat{\theta}_{BLUE,n}) = \mathcal{E}(\mu_n^*)$$

Construction of optimal n-point designs for BLUE is a difficult computational problem (Sacks, Ylvisaker (1965), etc.)

Comparison of $\mathrm{var}(\hat{\theta}_{OLSE,n})$ versus $\mathrm{var}(\hat{\theta}_{BLUE,n})$

Of course,

$$\begin{aligned}\mathrm{var}(\hat{\theta}_{BLUE,n}) &= 1/(1_n^\top K_n^{-1} 1_n) = \mathcal{E}(\mu_n^*) \leq \\ \mathrm{var}(\hat{\theta}_{OLSE,n}) &= \frac{1}{n^2}(1_n^\top K_n 1_n) = \mathcal{E}(\mu_n)\end{aligned}$$

Matrix analysis approach:

$$\frac{\mathrm{var}(\hat{\theta}_{OLSE,n})}{\mathrm{var}(\hat{\theta}_{BLUE,n})} = \frac{1}{n^2}(1_n^\top K_n 1_n)\,(1_n^\top K_n^{-1} 1_n) \geq 1$$

by the Cauchy-Schwarz inequality (as $1_n^\top 1 = n^2$), where we have equality if and only if 1_n is an eigenvector of K_n; that is, $K_n 1_n = \lambda 1_n$ for some $\lambda > 0$ ($=$ row sums of K_n are the same).

$\mathrm{var}(\hat{\theta}_{OLSE,n}) - \mathrm{var}(\hat{\theta}_{BLUE,n})$

$$u^\top u \leq (u^\top K_n u)\,(u^\top K_n^{-1} u) \leq \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2$$

Left inequality: Cauchy-Schwarz. Right inequality: Kantorovich (poor).

Better upper bound: using the optimality theorems, for a design $\mu (= \mu_n)$ and OLSE$= \int y(x)\mu(dx)$:

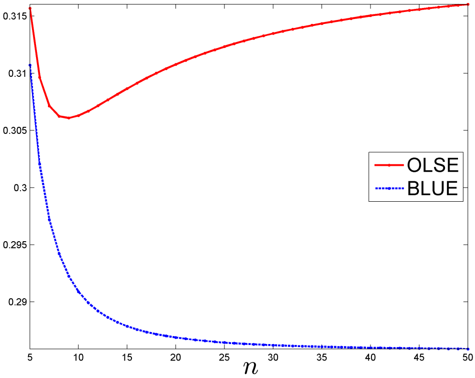
$$1 \geq \frac{\mathrm{var}(\hat{\theta}_{BLUE})}{\mathrm{var}(\hat{\theta}_{OLSE})} \geq 2\frac{\inf_{x\in\mathcal{X}} P_\mu(x)}{\mathcal{E}(\mu)} - 1$$

Useful relation:

$$\mathrm{MMD}^2(\mu_n, \mu_n^*) = \mathrm{var}(\hat{\theta}_{OLSE,n}) - \mathrm{var}(\hat{\theta}_{BLUE,n})\,.$$

Smit's paradox; behaviour of the variances as n increases

Smit's paradox (1961): for particular stationary kernels in one dimension and equidistant points, $\mathrm{var}(\hat{\theta}_{OLSE,n})$ increases for $n \geq n_0$.



Red curve converges to the energy of the limiting design (e.g. uniform). Where does the blue curve converge? How fast is the convergence? How different could be the two limits?

Q: What is the limit of the BLUE curve?

Let $x_1, x_2, \ldots$ be a dense sequence of distinct points in compact $\mathcal{X}$.
Designs: $X_n = \{x_1, \ldots, x_n\}$;
Model: $y(x_j) = \theta f(x_j) + \varepsilon(x_j)$ (we have $f(x) = 1_{\mathcal{X}}(x)$).

Theorem (Parzen, 1971)

$f \in H(K)$ if and only if

$$\mathrm{var}(\widehat{\theta}_{BLUE,n}) \to 1/\|f\|_{H(K)} = \mathrm{var}(\hat{\theta}_{BLUE,\infty}) \text{ as } n \to \infty.$$

Here $\hat{\theta}_{BLUE,\infty}$ is the continuous BLUE of θ but we may not be able to write it in the form $\hat{\theta}_{BLUE,\infty} = \int y(x)\mu(dx)$

$$\text{(recall } \hat{\theta}_{BLUE,n} = \int y(x)\mu_n^*(dx))$$

So, the limit of the BLUE curve is $1/\|1\|_{H(K)}$ but there may be no limit of the signed measures μ_n^* defining the discrete BLUEs.
Rate of convergence=MMD(μ_n^*, μ^*) if μ^* exists; optimal design for BLUE.

Q: When the limits of the two curves coincide?

Answer: When the minimum-energy probability measure is the minimum-energy signed measure μ^* (see above) and the limit of μ_n defining the OLSEs converges to this measure.

From any C(I)PD kernel K we can construct a kernel (reduced kernel) such that given μ is the minimum-energy measure:

$$K_\mu(x, x') = K(x, x') - P_\mu(x) - P_\mu(x') + \mathcal{E}(\mu)$$

A useful relation: $\mathrm{MMD}^2(\mu, \nu) = \mathcal{E}(\mu - \nu) = \mathcal{E}_{\mathcal{K}_\mu}(\nu)$.

Theorem (LP & AZ)

(i) $[K_\mu]_\nu = K_\nu$

(ii) *μ discrete: K CPD iff K_μ is PD*

(iii) *K is CIPD iff K_μ is IPD*

(iv) *μ has infinite support; K is bounded and CISPD $\Rightarrow K_\mu$ is ISPD*

(ii) is a generalization of Schoenberg's result in which μ is a delta measure. This property is very important, e.g., for the energy distance.

Special role of the constant function $1 = 1_{\mathcal{X}}(x)$

Questions:

(i) Does 1 belong to the RKHS $H(K)$?
(ii) Does 1 belong to the space of potentials $P(K)$?
$P(K) = \{P_\mu(x) = \int K(x, x')\mu(dx')$ for a finite signed measure $\mu\}$
(iii) In case of (ii), is μ a positive measure?

Importance:

(i) $\Leftrightarrow$ continuous BLUE exists
(ii) $\Leftrightarrow$ continuous BLUE exists and has the form $\hat{\theta}_{BLUE,\infty} = \int y(x)\mu(dx)$
(iii) $\Leftrightarrow$ continuous BLUE exists, has the form above and coincides with continuous OLSE for the optimal design.

More results: K is PD, $K(x, x') = \psi(x - x')$

- Spectral measure of ψ is moment-determinant (e.g., ψ is an analytic function) and has no mass at $0 \Rightarrow 1 \notin H(K)$ (H.Dette & AZ, 2021)
- If $\mathcal{X} = [0, 1]$, $\psi(t)$ is non-negative, non-constant, bounded and decreasing for $t > 0$, then the uniform measure μ_0 cannot be minimum-energy measure (simply $P_{\mu_0}(0.5) > P_{\mu_0}(0)$).

Conclusions

Conjectures; PD kernels $K(x, x') = \psi(x - x')$

1. $1 \notin H(K) \Rightarrow$ the spectral measure of ψ is moment-determinant or it has a positive mass at 0.
2. ψ is differentiable at $0 \Rightarrow 1 \notin P(K)$ (so that $\hat{\theta}_{BLUE,\infty} \neq \int y(x)\mu(dx)$ even if it exists)
3. Uniform measure on $\mathcal{X}$ cannot be minimum-energy (not true if $\psi(0) = \infty$)
4. Let $\psi(0) < \infty$, μ^* (if it exists) and μ^+ be minimum-energy measures ($\mu^*(\mathcal{X}) = \mu^+(\mathcal{X}) = 1$). Then $\mu^*(\partial\mathcal{X}) > 0$ and $\mu^+(\partial\mathcal{X}) > 0$.

Conjectures; PD kernels $K(x, x') = \psi(x - x')$

1. $1 \notin H(K) \Rightarrow$ the spectral measure of ψ is moment-determinant or it has a positive mass at 0.
2. ψ is differentiable at 0 $\Rightarrow 1 \notin P(K)$ (so that $\hat{\theta}_{BLUE,\infty} \neq \int y(x)\mu(dx)$ even if it exists)
3. Uniform measure on $\mathcal{X}$ cannot be minimum-energy (not true if $\psi(0) = \infty$)
4. Let $\psi(0) < \infty$, μ^* (if it exists) and μ^+ be minimum-energy measures ($\mu^*(\mathcal{X}) = \mu^+(\mathcal{X}) = 1$). Then $\mu^*(\partial\mathcal{X}) > 0$ and $\mu^+(\partial\mathcal{X}) > 0$.

Other directions of possible research ($K(x, x') = \psi(x - x')$, $d > 1$):

1. Extension of Smit's paradox to $d > 1$ and relaxation of conditions on ψ
2. Structure of minimum-energy measures
3.
4.