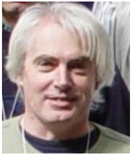
Optimal prediction measures

Joaquim Ortega-Cerdà

Universitat de Barcelona

June 3, 2021

Analysis of Monge-Ampère, a tribute to Ahmed Zeriahi

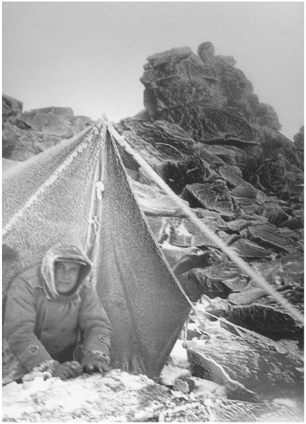
Joint work with Len Bos and Norm Levenberg

Gustav Elfving

The theory of optimal prediction tries to predict the value of a polynomial while only knowning the values of p at certain points in a restricted set K, maybe with an error. One tries to choose the points to evaluate p wisely.

Gustav Elfving

The theory of optimal prediction tries to predict the value of a polynomial while only knowning the values of p at certain points in a restricted set K, maybe with an error. One tries to choose the points to evaluate p wisely. One of the pioneers of the field was Gustav Elfving (1908-1984).



Optimal designs in polynomial prediction

Let $p \in \mathbb{C}_n[z]$ be a polynomial of degree n in $\mathbb{C}^d$

$$p = \sum_{k=1}^{N} \theta_k p_k$$

where $\{p_1, p_2, \ldots, p_N\}$ is a basis for $\mathbb{C}_n[z]$ and N its dimension.

Optimal designs in polynomial prediction

Let $p \in \mathbb{C}_n[z]$ be a polynomial of degree n in $\mathbb{C}^d$

$$p = \sum_{k=1}^{N} \theta_k p_k$$

where $\{p_1, p_2, \ldots, p_N\}$ is a basis for $\mathbb{C}_n[z]$ and N its dimension.
Let $K \subset \mathbb{C}^d$ and we observe the values of a particular $p \in \mathbb{C}_n[z]$ at a set of $m \geq N$ points $\{z_j \,:\, 1 \leq j \leq m\} \subset K$ with some random errors, i.e., we observe

$$y_j = p(z_j) + \epsilon_j, \quad 1 \leq j \leq m$$

where we assume that the errors $\epsilon_j \sim N(0, \sigma)$ are independent.

Optimal designs in polynomial prediction

Let $p \in \mathbb{C}_n[z]$ be a polynomial of degree n in $\mathbb{C}^d$

$$p = \sum_{k=1}^{N} \theta_k p_k$$

where $\{p_1, p_2, \dots, p_N\}$ is a basis for $\mathbb{C}_n[z]$ and N its dimension.
Let $K \subset \mathbb{C}^d$ and we observe the values of a particular $p \in \mathbb{C}_n[z]$ at a set of $m \geq N$ points $\{z_j \,:\, 1 \leq j \leq m\} \subset K$ with some random errors, i.e., we observe

$$y_j = p(z_j) + \epsilon_j, \quad 1 \leq j \leq m$$

where we assume that the errors $\epsilon_j \sim N(0, \sigma)$ are independent.
Let $z \in \mathbb{C}^d \setminus K$. We want to estimate $p(z)$ from the values y_j.

Prediction through least squares

First we want to estimate the parameters θ_k. The least squares estimate is provided by:

$$\widehat{\theta} := (V_n^* V_n)^{-1} V_n^* y.$$

where

$$V_n := \begin{bmatrix} p_1(z_1) & p_2(z_1) & \cdots & p_N(z_1) \\ p_1(z_2) & p_2(z_2) & \cdots & p_N(z_2) \\ \vdots & & & \vdots \\ p_1(z_m) & p_2(z_m) & \cdots & p_N(z_m) \end{bmatrix} \in \mathbb{C}^{m \times N}$$

is the Vandermonde matrix.

The variance of the prediction

The predicted value for $p(z)$ is given by

$$X = \sum_{k=1}^{N} \widehat{\theta}_k p_k(z)$$

which is a Gaussian random variable with variance:

$$\mathrm{var}(X) = \sigma^2 \mathbf{p}^*(z)(V_n^* V_n)^{-1} \mathbf{p}(z).$$

The variance of the prediction

The predicted value for $p(z)$ is given by

$$X = \sum_{k=1}^{N} \widehat{\theta}_k p_k(z)$$

which is a Gaussian random variable with variance:

$$\mathrm{var}(X) = \sigma^2 \mathbf{p}^*(z)(V_n^* V_n)^{-1}\mathbf{p}(z).$$

If $\mu_X = \frac{1}{m}\sum_{k=1}^m \delta_{x_k}$, then

$$\frac{1}{m} V_n^* V_n = G_n(\mu_X), \text{ where}$$

$$G_n(\mu) := \left[\int_K p_i(w)\overline{p_j(w)}d\mu(w)\right]_{1\le i,j\le N} \in \mathbb{C}^{N\times N}$$

is the Gram matrix of the polynomials p_i with respect to μ and

$$\mathrm{var}(X) = \frac{1}{m}\sigma^2 \mathbf{p}^*(z)(G_n(\mu_X))^{-1}\mathbf{p}(z)$$

And the Bergman kernel shows up

One checks that for any $\mu \in \mathcal{M}(K)$,

$$\mathbf{p}^*(z)(G_n(\mu))^{-1}\mathbf{p}(z) = K_n^\mu(z,z)$$

where, for $\{q_1, \cdots, q_N\} \subset \mathbb{C}_n[z]$, a μ-orthonormal basis for $\mathbb{C}_n[z]$,

$$K_n^\mu(w,z) := \sum_{k=1}^{N} \overline{q_k(w)} q_k(z)$$

is the reproducing Bergman kernel for $(\mathbb{C}_n[z], L^2(\mu))$. It satisfies

$$K_n^\mu(z,z) = \sup_{p \in \mathbb{C}_n[z]} \frac{|p(z)|^2}{\int_K |p(w)|^2 d\mu} = \sup_{p \in \mathbb{C}_n[z],\ p(z)=1} \frac{1}{\int_K |p(w)|^2 d\mu}.$$

The polynomial that achieves the sup is called a *prediction polynomial* for z with respect to μ.

Many optimal designs

Thus we are led to the following problem: Given $z \notin K$, minimize for all probability mesures $\mu \in \mathcal{M}(K)$ the Bergman kernel at the diagonal $K_n^\mu(z,z)$, find the measure that minimizes it and its corresponding prediction polynomial. This is the optimal design.

Many optimal designs

Thus we are led to the following problem: Given $z \notin K$, minimize for all probability mesures $\mu \in \mathcal{M}(K)$ the Bergman kernel at the diagonal $K_n^\mu(z,z)$, find the measure that minimizes it and its corresponding prediction polynomial. This is the optimal design.

There are other possible notions of optimal design. Most notably:

- G-optimal designs. We want to minimize

$$\min_{\mu \in \mathcal{M}(K)} \max_{z \in K} K_n^\mu(z,z)$$

- D-optimal designs. We want to maximize the determinant of the *design* matrix $G_n(\mu)$.

Many optimal designs

Thus we are led to the following problem: Given $z \notin K$, minimize for all probability mesures $\mu \in \mathcal{M}(K)$ the Bergman kernel at the diagonal $K_n^\mu(z,z)$, find the measure that minimizes it and its corresponding prediction polynomial. This is the optimal design.

There are other possible notions of optimal design. Most notably:

- G-optimal designs. We want to minimize

$$\min_{\mu\in\mathcal{M}(K)} \max_{z\in K} K_n^\mu(z,z)$$

- D-optimal designs. We want to maximize the determinant of the *design* matrix $G_n(\mu)$.

Kiefer and Wolfowitz have given a remarkable equivalence between both notions.

Some examples

The disk

Let $K = \{z \in \mathbb{C} : |z| \leq 1\}$ and $|z| > 1$. In this case one optimal prediction measure $\mu_0(z)$ is the Poisson kernel measure at the point $1/\bar{z}$ for all the degrees n.

Some examples

The disk

Let $K = \{z \in \mathbb{C} : |z| \leq 1\}$ and $|z| > 1$. In this case one optimal prediction measure $\mu_0(z)$ is the Poisson kernel measure at the point $1/\bar{z}$ for all the degrees n.

In such a case $K_n^{\mu_0}(z,z) = |z|^{2n}$. For any other probability measure μ supported on K

$$K_n^{\mu}(z,z) = \sup_{p(z)=1} \frac{1}{\int_K |p(w)|^2 \, d\mu(w)} \geq$$

$$\frac{1}{\int_K |\bar{z}^n w^n / |z|^{2n}|^2 \, d\mu(w)} = |z|^{2n} = K_n^{\mu_0}(z,z)$$

Some examples

The disk

Let $K = \{z \in \mathbb{C} : |z| \leq 1\}$ and $|z| > 1$. In this case one optimal prediction measure $\mu_0(z)$ is the Poisson kernel measure at the point $1/\bar{z}$ for all the degrees n.
In such a case $K_n^{\mu_0}(z,z) = |z|^{2n}$. For any other probability measure μ supported on K

$$K_n^{\mu}(z,z) = \sup_{p(z)=1} \frac{1}{\int_K |p(w)|^2 \, d\mu(w)} \geq$$

$$\frac{1}{\int_K |\bar{z}^n w^n / |z|^{2n}|^2 \, d\mu(w)} = |z|^{2n} = K_n^{\mu_0}(z,z)$$

In this case, for every n there are discrete measures (provided by Szego quadrature formulas for instance) that have the same n-moments as μ_0. Thus they are also optimal measures.

Let $K = [-1, 1]$. Then Hoel and Levine proved show that for *any* $z \in \mathbb{R} \setminus [-1, 1]$, a *real* external point, the optimal prediction measure is unique and is a discrete measure supported at the $n + 1$ extremal points $x_k = \cos(k\pi/n)$, $0 \leq k \leq n$, of $T_n(x)$ the classical Chebyshev polynomial of the first kind. In this case it turns out that

$$K_n^{\mu_0}(z, z) = T_n^2(z).$$

It is well known that T_n is a solution to many extremal problems.

Another extremal problem

For $K \subset \mathbb{C}^d$ compact and $z \in \mathbb{C}^d \backslash K$ an *external* point, we say that $P_n \in \mathbb{C}_n[z]$ has *extremal growth relative to K at z* if

$$P_n = \arg\max_{p \in \mathbb{C}_n[z]} \frac{|p(z)|}{\|p\|_K} \tag{1}$$

where $\|p\|_K$ denotes the sup-norm of p on K. Alternatively, we may normalize p to be 1 at the external point z and use

$$P_n = \arg\max_{p \in \mathbb{C}_n[z],\ p(z)=1} \frac{1}{\|p\|_K}.$$

Another extremal problem

For $K \subset \mathbb{C}^d$ compact and $z \in \mathbb{C}^d \backslash K$ an *external* point, we say that $P_n \in \mathbb{C}_n[z]$ has *extremal growth relative to K at z* if

$$P_n = \arg\max_{p \in \mathbb{C}_n[z]} \frac{|p(z)|}{\|p\|_K} \tag{1}$$

where $\|p\|_K$ denotes the sup-norm of p on K. Alternatively, we may normalize p to be 1 at the external point z and use

$$P_n = \arg\max_{p \in \mathbb{C}_n[z],\ p(z)=1} \frac{1}{\|p\|_K}.$$

$T_n(x)$ is the polynomial of extremal growth for any point $z \in \mathbb{R}\backslash[-1,1]$ relative to $K = [-1,1]$. Erdős has shown that the Chebyshev polynomial is also extreme relative to $[-1,1]$ for *real* polynomials at points $z \in \mathbb{C}$ with $|z| \geq 1$, i.e.,

$$\max_{p \in \mathbb{R}_n[x],\ \|p\|_{[-1,1]} \leq 1} |p(z)| = |T_n(z)|.$$

The connection

Proposition

The minimal variance is the square of the maximal polynomial growth, i.e.,

$$\min_{\mu \in \mathcal{M}(K)} K_n^\mu(z,z) = \max_{p \in \mathbb{C}_n[z],\, p(z)=1} \frac{1}{\|p\|_K^2}.$$

The connection

Proposition

The minimal variance is the square of the maximal polynomial growth, i.e.,

$$\min_{\mu\in\mathcal{M}(K)} K_n^\mu(z,z) = \max_{p\in\mathbb{C}_n[z],\, p(z)=1} \frac{1}{\|p\|_K^2}.$$

$$\min_{\mu\in\mathcal{M}(K)} \max_{p\in\mathbb{C}_n[z],\, p(z)=1} \frac{1}{\int_K |p(w)|^2 d\mu} =$$

$$= \frac{1}{\max_{\mu\in\mathcal{M}(K)} \min_{p\in\mathbb{C}_n[z],\, p(z)=1} \int_K |p(w)|^2 d\mu}.$$

The Minimax theorem

Now, for $\mu \in \mathcal{M}(K)$ and $p \in \mathbb{C}_n[z]$ such that $p(z) = 1$, let

$$f(\mu, p) := \int_K |p(w)|^2 d\mu.$$

f is linear in μ and convex in p.

The Minimax theorem

Now, for $\mu \in \mathcal{M}(K)$ and $p \in \mathbb{C}_n[z]$ such that $p(z) = 1$, let

$$f(\mu, p) := \int_K |p(w)|^2 d\mu.$$

f is linear in μ and convex in p.
By the Minimax Theorem

$$\max_{\mu \in \mathcal{M}(K)} \min_{p \in \mathbb{C}_n[z],\, p(z)=1} \int_K |p(w)|^2 d\mu = \min_{p \in \mathbb{C}_n[z],\, p(z)=1} \max_{\mu \in \mathcal{M}(K)} \int_K |p(w)|^2 d\mu.$$

and

$$\max_{\mu \in \mathcal{M}(K)} \int_K |p(w)|^2 d\mu = \|p\|_K^2.$$

A more precise version

It is also possible to give a more precise relation between the extremal polynomials for the two problems (of minimum variance and extremal growth).

Theorem

A measure $\mu_0 \in \mathcal{M}(K)$ is an optimal prediction measure for $z \notin K$ relative to K if and only if the associated prediction polynomial $P_n^{\mu_0,z}$ satisfies

$$\max_{w\in K} |P_n^{\mu_0,z}(w)|^2 = \int_K |P_n^{\mu_0,z}(w)|^2 d\mu_0,$$

or, equivalently, if and only the associated prediction polynomial is also a polynomial of extremal growth at z relative to K.

Back to the example $K = [-1, 1]$

The support of an optimal prediction measure in this case is a subset of $[-1, 1]$ where $|P_n^{\mu_0,z}(z)| = 1$, its maximum value. It has at most $2n$ points. They must be $n-1$ interior points and two end points ± 1. Consequently

$$\mu_0 = \sum_{i=0}^{n} w_i \delta_{x_i}$$

with weights $w_i > 0$, $\sum_{i=0}^{n} w_i = 1$.

Back to the example $K = [-1, 1]$

The support of an optimal prediction measure in this case is a subset of $[-1, 1]$ where $|P_n^{\mu_0,z}(z)| = 1$, its maximum value. It has at most $2n$ points. They must be $n-1$ interior points and two end points ± 1. Consequently

$$\mu_0 = \sum_{i=0}^{n} w_i \delta_{x_i}$$

with weights $w_i > 0$, $\sum_{i=0}^{n} w_i = 1$.

Proposition (Hoel-Levine)

Suppose that $-1 = x_0 < x_1 < \cdots < x_n = +1$ are given. Then among all discrete probability measures supported at these points, the measure with

$$w_i := \frac{|\ell_i(z)|}{\sum_{i=0}^{n} |\ell_i(z)|}, \quad 0 \le i \le n \tag{2}$$

with $\ell_i(w)$ the Lagrange polynomial, minimizes $K_n^{\mu}(z,z) = \left(\sum |\ell_i(z)|\right)^2$.

The case $z = ia$

In this case it is possible to compute the optimal prediction polynomial and optimal measure. The optimal polynomial in this case depends on the point a and satisfies a three term recurrence relation similar to the Chebyshev polynomials.

$$
\begin{aligned}
Q_1(z) &= -\frac{az+i}{\sqrt{a^2+1}}, \\
Q_2(z) &= \frac{1}{\sqrt{a^2+1}}\left(-(a+\sqrt{a^2+1})z^2 - iz + \sqrt{a^2+1}\right), \\
Q_{n+1}(z) &= 2zQ_n(z) - Q_{n-1}(z), \quad n = 2, 3, \cdots .
\end{aligned}
$$

The case $z = ia$

In this case it is possible to compute the optimal prediction polynomial and optimal measure. The optimal polynomial in this case depends on the point a and satisfies a three term recurrence relation similar to the Chebyshev polynomials.

$$\begin{aligned} Q_1(z) &= -\frac{az+i}{\sqrt{a^2+1}}, \\ Q_2(z) &= \frac{1}{\sqrt{a^2+1}}\left(-(a+\sqrt{a^2+1})z^2 - iz + \sqrt{a^2+1}\right), \\ Q_{n+1}(z) &= 2zQ_n(z) - Q_{n-1}(z), \quad n = 2, 3, \cdots . \end{aligned}$$

Theorem (Bos)

The optimal prediction measures $\mu_n \overset{*}{\rightharpoonup} \mu$ *which is the push forward of the Poisson measure by the Joukowski map* $J(z) = \frac{1}{2}(z + 1/z)$ *at the point* $1/J^{-1}(-ia)$.
Moreover μ_n *are the Gauss-Lobato measures associated to* μ *of degree* n.