Mechanical models in nonparametric regression

Vladimír Balek\textsuperscript{1,*} and Ivan Mizera\textsuperscript{2,†}

Comenius University and University of Alberta

A common motif in the expositions of spline-based methods in statistical smoothing or numerical interpolation is an allusion to mechanical analogies—motivated perhaps by a desire to provide some explanation why the resulting shapes ought to be regarded as “natural”. The univariate case has its Oxford English Dictionary reference to draftsman spline as “a flexible strip of wood or hard rubber used by draftsmen in laying out broad sweeping curves”, which suggests (amiss!) that the eponymous mathematical object shares exactly the same properties. The introduction of “thin-plate spline” in the bivariate domain usually comes with a more distinctive story about the deformation of an elastic flat thin plate—for instance, page 139 of Green and Silverman [14] or page 108 of Small [32]: if the plate is deformed to the shape of the function $\epsilon f$, and $\epsilon$ is small, then the bending energy is (up to the first order) proportional to the smoothing penalty.

The importance attached by the scientific community to such trivia varies: while some consider it a signal from Nature, indicating the righteous path in the potentially endless forest of possibilities—see especially Bookstein [3, 4, 5], but also Bookstein and Green [6], Small [32], Dryden and Mardia [10]—for others it is a marginal curiosity, not deserving to stand in the path of the appreciation of computational and theoretical properties.

In our case, the desire of the second author to comprehend the connection between thin-plate splines and total variation penalties led him to cross-questioning of the first author, theoretical physicist with principal interests in gravitation and cosmology; the latter reluctantly, but eventually cooperatively descended into the caverns of “engineering”—theory of elastic and plastic behavior of solid bodies. The unveiled connection not only turned out to be interesting, but yielded also a practical return for the second author: the elucidated mechanical models hinted Koenker and Mizera [19] where—that is, in which community—to look for relevant algorithmic solutions for their proposals.

Which could be the end of the story were it not for queries that started to come occasionally thereafter, about a text documenting the apocryphal knowledge. So, here finally an attempt to produce one. The available space allows only for an overview of physical facts relevant to nonparametric regression; for somewhat nonstandard physical derivations (albeit perhaps unsurprising for an expert in solid mechanics), we refer to Balek [2].

\textsuperscript{*}Research supported by the grant VEGA 1/1008/09.
\textsuperscript{†}Research supported by the NSERC of Canada.

\textsuperscript{1}Department of Theoretical Physics, Comenius University, Mlynská dolina, Bratislava, SK-84248, Slovakia.

\textsuperscript{2}Department of Mathematical and Statistical Sciences, CAB 632, University of Alberta, Edmonton, AB, T6G 2G1, Canada.
1. Regularization in nonparametric regression

Regularization, or “roughness penalty approach” seeks a nonparametric regression fit $\hat{f}$ via minimizing

\begin{equation}
L(z, f) + J(f) \rightarrow \min_f \tag{1}
\end{equation}

The usual formulation features $\lambda J(f)$ instead of $J(f)$, to facilitate the tuning of the influence of $J$ through $\lambda > 0$. As this aspect will not be essential here, we use rather the expression (1), with $\lambda$ considered as included in $J$, if necessary.

The first part of (1), $L$, is a certain lack-of-fit, “loss” function. It depends on the components, $z_i$, of the response vector $z$; and also on the fitted function, $f$, but only through its values at predictor points $u_i$. An example is

\[ L(z, f) = \sum_{i=1}^{n} (z_i - f(u_i))^2. \]

Another, somewhat artificial example may be conceived to express numerical interpolation in the framework of (1) : $L(z, f) = 0$ when all $z_i = f(u_i)$, otherwise $L(z, f) = +\infty$. This makes the minimization of (1) equivalent to

\begin{equation}
J(f) \rightarrow \min_f \text{ subject to } z_i = f(u_i) \tag{2}
\end{equation}

The second part of the objective function, $J$, is traditionally called penalty. It is responsible for the shape of $\hat{f}$, the configuration the minimizing $f$ will take. Indeed, if the process of solving (1) is divided into two stages, and in the first stage the fitted values, $\hat{f}(u_i)$, are determined, then the second stage amounts to solving the interpolation problem (2) with $z_i$ replaced by $\hat{f}(u_i)$, as the objective in (1) then depends only on $J(f)$.

2. Penalty as deformation energy

A typical penalty designed with an intent to make the resulting fit smooth is set to control some derivative(s) of its argument; in the univariate case, the typical instance is

\begin{equation}
J(f) = \int_{\Omega} (f^{(k)}(x))^2 dx = \int_{\Omega} (f^{(k)})^2 dx. \tag{3}
\end{equation}

Hereafter, we will omit functional arguments in the penalty expressions as indicated by the second equality in (3)—the convention that will save us considerable amount of space.

The square in the penalty (3) implies that the sign of the derivative is not essential, only its magnitude; the preference to, say, absolute value may be dictated by computational convenience. The integral gives a summary of the behavior of the derivative; for now, we postpone the discussion of the aspects of the domain of integration $\Omega$.

It may be not entirely clear which order of derivative, $k$, is to be chosen, and there are very few guidelines regarding this. One of the possible ones is to look at the limiting behavior at zero: it suggests that the fits are shrunk toward $f$ with $J(f) = 0$. That is, toward constants if $k = 1$, toward linear functions if $k = 2$, toward quadratic if $k = 3$. 

The physical model singles out $k = 2$, yielding the celebrated cubic spline penalty

$$J(f) = \int_{\Omega} (f'')^2 \, dx.$$  

This model envisions the graph of $f$ as a solid body deformed from the initial relaxed, zero energy configuration $f \equiv 0$, and considers the energy of this deformation, the work that had to be done to transform $f$ from the relaxed configuration to the present one. It is then attempted to identify this energy with (some) $J(f)$; if the deformation was caused by the action of external forces at points $u_i$, resulting in a configuration of $f$ satisfying $z_i = f(u_i)$, then the interpolation problem (2) can be interpreted as a physical variational principle stipulating that the resulting shape is the one with minimal energy, achieved with minimal effort.

And indeed, if the deformation is elastic—which means that when the forces are lifted, then $f$ assumes back its relaxed configuration—then the cubic spline penalty (4) is equal to the first-order approximation of the energy of this deformation, approaching the true value under certain limiting conditions.

However, if the deformation is irreversible, plastic, then under similar conditions the first-order approximation of the deformation energy leads to a penalty

$$J(f) = \int_{\Omega} |f''| \, dx = \sqrt{\int_{\Omega} f'},$$

the total variation of $f'$ on $\Omega$. This is the penalty used by Koenker, Ng and Portnoy [21] in their proposal for “quantile smoothing splines” and further cultivated by He, Ng and Portnoy [16] (the equality of the integral to the total variation is the Vitali theorem, in its classical version constricted to absolutely continuous functions; the use of derivatives in the sense of Schwartz distributions enables full generality). The penalty (5) is a member, for $k = 2$, of the family analogous to (3),

$$J(f) = \int_{\Omega} |f^{(k)}| \, dx = \sqrt{\int_{\Omega} f^{(k-1)}},$$

considered by Mammen and van de Geer [24], Koenker and Mizera [19, 20], and others.

The bivariate case brings more possibilities, but the overall picture is about the same: elastic deformation energy involves square, or squared norm; plastic energy absolute value, or norm. The relationship of quadratic penalties to energies of elastic deformations is well known, and even a non-expert can easily locate the relevant expressions in the standard physics textbooks like Landau and Lifshitz [22]. There are certain details one can learn in this process, and they are definitely of some interest, but probably would not warrant an autonomous publication. On the other hand, it seems that the connection of total variation penalties to deformation energy in plasticity may not be that obvious; in particular, the relevant variational principles seem not to be available in the existing physical or engineering literature on plasticity theory. This does not mean that such a development would require non-trivial theoretical breakthroughs—but rather, as we note below, that practical theory of plasticity is concerned with different types of problems.

As the relevant variational principles in plasticity theory bear some analogy to those in elasticity, we start by review of the latter.
3. Elastic deformation

The central notion used in calculating the deformation energy $J(f)$ is stress, the distribution of internal tension forces caused by the external forces; deformation energy is the work done by the stress in the course of deformation; energy density (energy per unit volume) can be obtained via integration involving locally described stress and deformation, the quantities linked by so-called constitutive equation.

We have defined elastic deformations as the reversible ones. They occur only if the load does not exceed certain critical value. For most of materials, the constitutive equation in the bulk of the interval of admissible loads is the (generalized) Hooke law, stipulating that stress is proportional to deformation. A special case is the classical, one-dimensional Hooke law (according to which the relative dilation is proportional to the pressure load), by applying it to an infinitesimal three-dimensional element. The relationship between stress and deformation is thus linear, but only under the condition (underlying both Hooke laws) that the deformation is small; it abides by a principle that small causes yield linear responses.

The physical model in the univariate case is that of a solid rod, whose shape (more precisely, the shape of its axis, if the width is not neglected) is given by $f$. The Hooke law is applied to the stretching of layers of the rod (the upper ones are stretched, the lower compressed); while the mean dilation is zero, the mean squared dilation is not, and the geometry of the problem suggests that the latter is proportional to the square of the curvature of the rod axis—which can be for small deformations approximated by $(f'')^2$. The energy density is proportional to the dilation squared and its integration over the cross section of the rod and subsequently over the length yields the equality of the total deformation energy to (4), up to a multiplicative constant that can be made 1 by the choice of units.

In the bivariate case, the role of the rod is taken on by a plate. The additional feature that needs to be taken into account now is that stretching typically causes squeezing in the perpendicular direction; the size of this effect is measured by an important characteristic of materials, the Poisson constant (ratio) $\nu$. If $\nu = 0$, there is no lateral squeezing, and the bending of the plate can be viewed as univariate stretching/squeezing in the directions of principal axes, the directions of the eigenvectors of the Hessian $H$. Denoting the corresponding eigenvalues by
Mechanical models in nonparametric regression

$h_1, h_2$, we obtain the deformation energy as the integral of $\text{tr} \, H^2 = h_1^2 + h_2^2$; it yields the celebrated thin-plate spline penalty (the subscripts indicate partial derivatives)

$$J(f) = \int_{\Omega} \left( f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \right) \, dx \, dy,$$

see Green and Silverman [14], Wahba [35]. If $\nu$ is nonzero, then the stress acting in one principal direction affects the curvature also in the other one, resulting in the additional contribution to the energy amounting to $2\nu$ times the integral of $\det H = h_1 h_2 - f_{xx} f_{yy}$. The deformation energy is then

$$J(f) = \int_{\Omega} \left( f_{xx}^2 + 2(1 - \nu)f_{xy}^2 + f_{yy}^2 + 2\nu f_{xx} f_{yy} \right) \, dx \, dy.$$

An alternative way to arrive to this expression is via rotational invariance arguments, in the spirit of Landau and Lifshitz [22]: the fact that a physical law for an isotropic body should not depend on a coordinate system implies that the integrand of (8) should be a bilinear, orthogonally invariant function of $H$; if the energy is presumed to be a quadratic function, such invariants are only $\text{tr} H^2$ and $\det H$. The assumptions here are that the plate is thin (its thickness being much smaller than the characteristic scale of deformation) and its stretching is negligible compared to bending—more precisely, no stretching forces are applied at the edge of the plate, and the bending of the plate is small (the deflection being much smaller than the thickness of the plate).

In the smoothing context, we are interested in instances of (8) that promise simple numerical treatment. One of them is (7); for $\nu = 1$, we obtain the squared Laplacian penalty

$$J(f) = \int_{\Omega} (f_{xx} + f_{yy})^2 \, dx \, dy = \int_{\Omega} (\Delta f)^2 \, dx \, dy,$$

which can be found in the works of O’Sullivan [27], Duchamp and Stuetzle [11], Ramsay [28], and others. It could be also called “the penalty of Sophie Germain”, because it emerged in her first attempt to win the prize of the French Academy of Sciences (in the contest initiated by Napoleon). While the condition $-1 \leq \nu \leq 1$
follows by the simple requirement of positive definiteness of (8) (ensuring that the
deformation energy works well in physical variational principles and the smoothing
penalty yields a convex optimization problem), a three-dimensional analysis shows
that \( \nu \) must not exceed \( \frac{1}{2} \). Hence the value \( \nu = 1 \) is “unphysical”.

Sophie’s choice is understandable in the light of the fact that the underlying
differential equation (corrected by a supportive referee Lagrange) does not involve \( \nu \).
In the presence of unsupportive members like Poisson (whose 1814 memoir on
plates, elaborating on the second, corrected version of Germain’s work, is considered
to be the first step toward general elasticity theory), the jury refused to award her
the coveted prize (despite the fact that she was the only contestant) even on the
second attempt, citing unclear physical justification. Only the third attempt two
years later brought her the pledged kilogram of gold.

Another distinctive value, \( \nu = -1 \) (materials with \( \nu < 0 \), expanding when they
are stretched, were discovered only recently), yields a penalty

\[
J(f) = \int_\Omega \left( (f_{xx} - f_{yy})^2 + 4f_{xy}^2 \right) \, dx \, dy,
\]

which seems not yet encountered in a smoothing context. Neither do the instances
of (8) with general \( \nu \), from the optimization point of view convex combinations of
penalties (7), (9), and (10). The substance with the Poisson constant \( \nu \) close to
0 is cork—it indeed exhibits almost no lateral expansion under compression. For
typical materials, the values of \( \nu \) are between 0.3 and 0.4 (steel 0.30, copper 0.34,
gold 0.42).

4. Plastic penalties

As already mentioned, the difference between elasticity and plasticity is reversibility.
Typically, increasing the forces applied to a body first cause an elastic deformation
(unless we face the instance of rigid plasticity, a rather idealized behavior) until they
reach a critical value called yield limit, after which the transition to plastic behavior
occurs. A theory that describes this behavior in its initial stages in a way analogous
to elasticity theory is called the deformation theory of plasticity. It relates stress
to the deformation in a similar way in the plastic as in the elastic regime, viewing
plasticity as a kind of nonlinear elasticity. It is this theory (more precisely, its version
for perfectly plastic rigid-plastic bodies) that provides a physical model for certain
alternative penalties in nonparametric regression. The theory can be extended to
include so-called hardening (whose presence, as opposed of the idealized behavior
of perfect plasticity, means that further plastic deformation requires increasingly
higher stresses). However, the principal drawback of any deformation theory is that
it does not account for irreversibility—this motivated other theories of plasticity,
notably incremental, or flow plasticity theory; see Kachanov [18]. It may be of
interest that in spite of all simplifications, the deformation theory of perfect rigid
plasticity is extensively applied in engineering, as the basis of limit analysis—the
study of limit loads, when plasticity takes over, and the subsequent deformations;
in other words, the theory helps to figure out when structures like roofs and bridges
can collapse.

The starting notion of this theory is a yield criterion, in two dimensions represented
by an orthogonal-similarity-invariant norm \( \| \cdot \| \) on \( 2 \times 2 \) symmetric matrices
involved in the constitutive inequality

\[
\| M \| \leq \mu,
\]
where $M$ is the moment matrix (related to stress) and $\mu$ indicates the threshold value: the plate is deformed elastically or remains flat if $\|M\| < \mu$, and may be deformed plastically if $\|M\| = \mu$. The latter equation is used for common visualization of the yield criterion in the space of eigenvalues, the yield surface. The constitutive equation is then obtained from the fact that the Hessian, $H$, is proportional (with negative proportionality constant) to the gradient of $\|M\|$. The ensuing deformation energy is

$$
\int_\Omega \|H\|_\ast \, dx \, dy,
$$

where $\| \cdot \|_\ast$ is a norm dual (conjugate) to $\| \cdot \|$; see Collatz [8]. This is in agreement with the general concept of duality in the mechanics of deformable bodies, discussed in detail in Témam [33].

\[\text{Various yield criteria}\]

The plastic properties of a particular material are thus characterized not only by a yield limit $m_0$, but also by a yield criterion. There are quite a few of these in the literature, designed to capture specific properties of various materials. One of the first dates back to 1868, and was proposed by Tresca, an engineer involved in the establishment of the international meter standard as we know it today. Following an abandoned idea by Maxwell from 1863, Richard von Mises proposed the elliptic criterion that bears his name, sometimes also connected with that of Maximilian Huber, who published some preliminary ideas in his article in Polish. While the original objective was to simplify the analysis, it turned out that von Mises criterion did even better when confronted with experimental data.

From our perspective, a family of yield criteria parametrized by a dimensionless parameter $\kappa \in [0, 1/2]$ (a “plastic Poisson constant”), proposed by Yang [37] with an objective to achieve better description of materials like marble and sandstone, gives similar expressions for deformation energy as in the elastic case,

$$
J(f) = \int_\Omega \sqrt{f_{xx}^2 + 2(1-\kappa)f_{xy}^2 + f_{yy}^2 + 2\kappa f_{xx}f_{yy}} \, dx \, dy.
$$

The family contains von Mises yield criterion for $\kappa = 1/2$. An example of another criterion with a simple mathematical expression is the square criterion deemed appropriate for concrete plates; see Mansfield [25]; it corresponds to the $\ell^\infty$ and its dual to the $\ell^1$ norm.
5. The domain of integration and other aspects

The recognition of the potential of elastic considerations in interpolation and smoothing dates back at least to Sobolev in 1950’s. In the variational problem with a penalty like (7), the functional is quadratic—the Euler-Lagrange equation is thus linear, the solutions have the superposition property, they form a linear space, and for fixed values of covariates \((x_i, y_i)\) a finite-dimensional one. The linear space is a Hilbert space, which opens a possibility to construct solutions via reproductive kernels, introduced in the smoothing context by Wahba [35] and made immensely popular by the recent statistical machine learning literature, where this particular approach opens tremendous range of new possibilities; in this context, the penalties are often eclipsed by kernels, from which the properties of the resulting solutions are inferred rather in an intuitive fashion; see Schölkopf and Smola [30].

From the physical point of view, the domain of integration \(\Omega\) is naturally determined by the solid body represented by \(f\); consequently, it naturally comes as bounded. From the smoothing perspective, however, the need to choose the domain may be an undesirable detail; this nuisance was overcome by the idealization idea of Harder and Desmarais [15], perfected by Duchon [12] and Meinguet [26]; see again Wahba [35]. It turns out that it is possible to set \(\Omega\) to be the entire \(\mathbb{R}^2\); it is then possible to obtain expressions for the basic functions in closed form, which substantially facilitates the application of Hilbert-space methods, and reduces the whole problem to a system of linear equations.

The resulting technology became a standard part of numerical and statistical methodology—usually referred to as “thin-plate splines”, where the idealization \(\Omega = \mathbb{R}^2\) is assumed implicitly, and special adjectives are added rather otherwise. For example, Green and Silverman [14] speak about “finite-window thin-plate splines” in case when \(\Omega\) is bounded; their interest in such a variant is motivated by the desire to achieve better behavior near the boundary of the data cloud (which may be undesirably influenced by the fact that the domain is unbounded).

The introduction of total variation penalties (6) by Koenker, Ng and Portnoy [21] in nonparametric regression, and by Rudin, Osher and Fatemi [29] in image reconstruction predated the recent interest in regularization with \(\ell^1\) penalties, spawned by Chen, Donoho and Saunders [7] and Tibshirani [34]. The appeal of these penal-
ties lies in their possibility to achieve sparse fits, the fits with many zero coefficients; for the total variation penalties in nonparametric regression context this translates to the possibility of capturing qualitative features like spikes and edges, without oversmoothing the relevant information; see Davies and Kovac [9] for an overview of this aspect. In the $\ell^1$ context, however, the relevant linear spaces of solutions are no longer Hilbert spaces; the possibility of using reproducive kernels is lost, and with it possible one potential attraction of unbounded integration domains. Nevertheless, the inherent sparsity helps in application of efficient methods of convex optimization, like interior point algorithms.

Even in the $\ell^2$ context, solving linear equations may not be that straightforward anymore when the system is large—unless its matrix is sparse, exhibiting relatively only few nonzero entries. The matrices arising in closed-form solutions obtained by reproducive kernels are non-sparse—and there does not seem to be any workaround for this; see Silverman [31], Wood [36]. A possible way out is to trade the “exact” solution for an approximate one, but with a sparse linear system involved in its computation; which is the case, for instance, in the finite-element implementations by Hegland, Roberts and Altas [17], Duchamp and Stuetzle [11], Ramsay [28], and others. The algorithms thus return back to that of boundary value problems, where the idealization of $\Omega$ to $\mathbb{R}^2$ is rather a burden, and in no case any longer a necessity.

6. Elastica

It is important to emphasize once again that the formula (4) expresses only “first-order”, “linearized” energy of a bent rod; the underlying assumption of the deflection being small should be kept in mind. If this is not the case, and the bending is not that small, then it is not surprising that mathematical splines (those based on the linearized energy expression) fail to meet the expectations set up by their Oxford English Dictionary eponyms—as in the picture showing a situation not that rare in practical interpolation.

\[
\text{Not flexible enough}
\]

One would like to have an interpolant preserving the monotonicity apparent in the data; this is not the case with the standard cubic spline.

Apparently, the deformation in this case is not “small”. While in more complex situation resorting to the leading (“first-order”) term may be the only way how to achieve some tractability, the case of elastic rod is simple enough to allow for exact solution even for large deflections: the deformation energy of the Euler elastica is, up to a constant, equal to

\[
(14) \quad J(f) = \int \frac{(f''')^2}{(1 + (f')^2)^{5/2}} \, dx.
\]

Compared to (4), we note that the denominator vanishes if $f'$ approaches zero; but also we note that $J$ is not convex in $f$, and thus one may expect problems
with existence and uniqueness of solutions. And indeed, Malcolm [23] mentions it as an open problem, while proposing methods how to obtain solutions for “non-linear splines”. The desire of achieving better monotonicity properties than those exhibited by cubic splines led to several proposals in the interpolation literature (one of them implemented as the MATLAB interpolation default); elastica, due to its computational difficulty, serves there rather as an ideal standard, by which all (allegedly simpler) proposals are gauged. In statistics, it seems that a need for this line of development has not been felt yet.

7. An archetypal example: radial interpolation

To get a better understanding for possible differences between various penalties deriving from mechanics of plates, it is instructive to solve the simplest problem in this setting, a circular plate lifted at the center. This is a classical exercise, and was worked out for an elastic simply supported plate by Poisson in 1829. Let us suppose that the plate is fixed at the radius 1, that is, \( f \) is equal to 0 at the boundary of the unit disk, and the center of the plate is lifted to 1, that is, \( f \) is equal to 1 at the origin. Strictly spoken, this is not a problem corresponding to point interpolation, but can be thought of as an idealization of a problem when \( f \) interpolates 1 at origin, and 0 at many points scattered in some uniform manner along the boundary. Recall also that the theory stipulates the deflection being much smaller than the typical dimension of the plate; therefore physically sensible solutions require rescaling by small \( \epsilon \), working with \( \epsilon f \) instead of \( f \).

The symmetry of the variational problem implies that \( f(x, y) = w(r) \), where \( r = (x^2 + y^2)^{1/2} \). For elastic penalties (8), we obtain after changing to polar coordinates

\[
J(f) = 2\pi \int_0^1 (r w'')^2 + \frac{(w')^2}{r} + 2\nu w'' w' \, dr.
\]

The Euler-Lagrange equation for minimizing \( J(f) \) subject to the conditions mentioned above is the biharmonic (Lagrange) equation, expressed through the radial part of the Laplace-Beltrami operator. The general form of \( w \) satisfying this equation is

\[
w(r) = Ar^2 \log r + Br^2 + C \log r + D.
\]

The boundary condition \( w(0) = 1 \) immediately yields \( C = 0, D = 1 \). After this simplification, we may either use physical reasoning, observing that the plate in

\[\text{Elastic penalties on a disc}\]

Left panel shows the sections of solutions for \( \nu = -1, 0, 0.42, 1 \), all with domain of integration being the disc with \( R = 1 \); the right panel considers \( \nu = 0 \), and \( R = 1, 1.5, \infty \).
this setting is simply supported at the edge, which entails the boundary condition
\begin{equation}
\label{16}
w(1) = w''(1) + \nu w'(1) = 0,
\end{equation}
or we may evaluate the penalty $J(f)$ for the general form (15) of $w$ and find the coefficients giving the minimum. Any of these ways leads to
\begin{equation}
w(r) = r^2(A \log r + B) + 1 = r^2\left(\frac{2(1+\nu)}{3+\nu} \log r - 1\right) + 1
\end{equation}
and the optimized value of the penalty
\begin{equation}
J(f) = 4\pi A = \frac{8\pi(1+\nu)}{3+\nu}
\end{equation}
— all when the integration domain is the unit disc, the disc with radius $R = 1$; we can see the plot of $w$ for various $\nu$ in the left panel of the picture. From the smoothing perspective, we may feel that (inspired by the piecewise linear interpolation featured in the previous section) that the shape underlying the data is the cone centered at the origin; while we perhaps want the final solution to be smooth, we also do not want it to yield too much to the possible rigidity of the smoothing scheme. In other words, we may have the desire for having the shape, while smooth, as sharp as possible at 0. This would deter us from $\nu = -1$, which in fact interpolates by a quadratic surface, and drive us rather to larger $\nu$, ultimately to $\nu = 1$. However, it has to be said that shapes for various $\nu$ differ only very little, so eventually we may well choose $\nu$ rather on the basis of computational convenience—which may favor, in particular, $\nu = 0$.

If we increase the integration domain, for simplicity still taking the disc, but with radius $R > 1$, the shape of the plate becomes different. For $\nu = 0$, using the similar methods as for $R = 1$, we obtain
\begin{equation}
w(r) = \begin{cases} r^2 (p \log r - 1) + 1, & \text{when } r \leq 1, \\ -p \log r + (p - 1)(R^2 - 1), & \text{when } 1 < r \leq R, \end{cases}
\end{equation}
where $p = (1 - 2R^{-2})^{-1}$. The optimized value of the penalty is $4\pi p$. We can see the resulting shapes in the right panel of the picture; the tendency with increasing $R$ is similar to that when $\nu$ is increased. Letting $R \to \infty$, we obtain that $p \to 1$, which gives the celebrated thin-plate spline solution of Harder and Desmarais [15] and Duchon [12] for $R = \infty$,
\begin{equation}
w(r) = \begin{cases} r^2 (\log r - 1) + 1, & \text{when } r \leq 1, \\ -\log r, & \text{when } r > 1, \end{cases}
\end{equation}
with the optimized penalty $4\pi$. Note that the solution inside the unit disc is equal to that with $\nu = 1$ and $R = 1$, however, it can be shown (via the Green theorem) that the solution, as whole, does not depend on $\nu$ in this case. Radial functions emerging here play an important role in finding the closed-form interpolating and smoothing fits via the reproducing-kernel Hilbert space theory.

For plastic penalties with general norms, the solutions are again radial functions; however, the corresponding sections have to be generally computed numerically. Nevertheless, the special case of (13) with $\kappa = 0$ can be solved in closed form, again for the radial integration domains, discs with radius $R$. The easiest case is
\( R = 1 \), when the solution can be obtained by guessing (and subsequently proving):

\[ w(r) = 1 - r, \text{ for } r \text{ going from } 0 \text{ to } 1. \]

This is an instance of a more general setting: for any \( R \) in the interval \([1, R_{\text{crit}}]\), there is a unique \( q \in [0, 1) \) such that

\[ w(r) = 1 - r^{1-q} \quad \text{for } 0 \leq r \leq 1. \]

The formula for extending these solutions to \( r \) ranging from 1 to \( R \) is too complicated to be given here: we only show a picture for values of \( q = 0.4, 0.8, 0.9 \), corresponding approximately to \( R = 1.375, 1.660, 1.715 \), and note the tendency of the solutions—which is not asymptotic: for \( R \geq R_{\text{crit}} \), the critical value approximately equal to 1.765, the solution degenerates: it is equal to 1 at 0 and to 0 elsewhere.

The solution of the interpolation problem for the plastic penalty (with \( \kappa = 0 \)) on a disc with radius \( R \) equal to 1, 1.375, 1.660, 1.715.

8. Blends

The last paragraph brings us to, in our opinion not very well understood phenomenon of increasing need of derivatives in regularization problems, related in some sense to the embedding theory of functional spaces of Sobolev type: certain penalties may not yield nondegenerate solutions. The radial interpolation example, interpolating 1 in the center of the disc and 0 at its boundary, is in a sense a test bed in this respect; compare, for instance, page 160 of Green and Silverman [14]. As we have no ambition here to go into higher-dimensional theories, let us mention only the example fairly well known: if instead of some penalty using second derivatives we take the Dirichlet penalty

\[
J(f) = \int_{\Omega} \left( f_x^2 + f_y^2 \right) \, dx \, dy,
\]

in physics corresponding not to elastic or plastic deformation of a non-stretching thin plate, but instead to the deformation of a membrane, an elastic plate where stretching is dominant. The solution of the radial interpolation is then degenerate—more precisely, it does not exist.

A common solution here is then to regularize the problem by considering a (realistic) situation when there is some stretching and some bending; in other words, consider a “blended” penalty, a convex combination of (17) and, say, (7). Another motivation for doing this may stem from the desire to alleviate the rigidity of thin-plate spline solution, as hinted in the previous section: including “tension” in the
spline prescription yields a pronounced peak in radial interpolation. The picture shows the result of the radial interpolation for such convex combination, depending on $\alpha$, where the coefficient at the membrane term (17) is $\alpha^2/(1 + \alpha^2)$; the broken lines correspond to the simplified version, the thin-plate splines with tension of Franke [13].

![Thin-plate penalties with tension](image)

The result of radial interpolation with convex combination of penalties, for $\alpha = 1/2, 2, 4, 10$; broken lines are thin-plate spline with tension of Franke [13].

In a similar manner, one can blend elastic and plastic penalties into “elasto-plastic” ones; however, the price paid for increased flexibility of interpolation and smoothing solutions, is the necessity of dealing with additional tuning parameters.

References


