IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 5–19

DOI: 10.1214/12-IMSCOLL902

Mechanical models in nonparametric regression

Vladimír Balek[1,*] **and Ivan Mizera**[2,†]

Comenius University and University of Alberta

This note is devoted to an extrapolation of a solid-mechanic motif from nonparametric regression (or, as it is often referred to, smoothing) and numerical interpolation, a theme originating in the exposition of spline-based methods. The attention is turned to the fact that a mechanical analogy, akin to that from elasticity theory for quadratic spline penalties, can be elucidated in plasticity theory for certain penalties of different type. It is this—in the (slightly altered) words of a referee, "what would happen if the metaphor underlying splines were replaced by a metaphor of plastic energy"—what is the objective of what follows; everything else, in particular discussion of the origins of (certain types of) splines, or an exhaustive bibliographical account, is outside our scope.

To give an idea what mechanical analogies we have in mind, let us start by the quote from the Oxford English Dictionary, referencing (univariate) "draftsman spline" as *"a flexible strip of wood or hard rubber used by draftsmen in laying out broad sweeping curves"*. Bivariate case brings more sophistication: the designation "thin-plate spline" is regularly explained by a story about the deformation of an elastic flat thin plate—see page 139 of Green and Silverman [12] or page 108 of Small [31]: *if the plate is deformed to the shape of the function ϵf, and ϵ is small, then the bending energy is (up to the first order) proportional to the smoothing penalty.* While it is questionable how much of scientific utility such trivia may carry—for some discussion, see Bookstein [3], Dryden and Mardia [7], Wahba [35], Green and Silverman [12], and the references there—their mere existence may provoke inquiring minds. While a curious individual may pick up the "physics" of the elastic parable pretty much from the standard textbooks (either of engineering [34] or theoretical flavor [21]), comprehension of the pertinent plastic parallel is much more perplexing; even if the topic is perhaps not unknown to the specialized literature, the latter may be quite impenetrable for a person with standard statistical education. Fortunately, with a little help from his friend (the first author of this note—a theoretical physicist, albeit with principal interests in gravitation and cosmology), the second author was not only able to arrive to certain joy of cognition, but also to

*Research supported by the grant VEGA 1/1008/09.

†Research supported by the NSERC of Canada.

‡We gratefully acknowledge the use of the illustration "The Greatest Mechanick this day in the world. Portrait of Robert Hooke, detail from the frontispiece of Chamber's Cyclopaedia, 1728 ed.", downloaded from www.she-philosopher.com.

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

2Department of Mathematical and Statistical Sciences, CAB 632, University of Alberta, Edmonton, AB, T6G 2G1, Canada.

a practical outcome: the elucidated mechanical model hinted Koenker and Mizera [17] where to look for relevant algorithmic solutions for their problems.

Given the available space and the anticipated general level of the interest, we limit ourselves here merely to an overview of the relevant physical facts and their implications in nonparametric regression. The somewhat nonstandard (but perhaps unsurprising for an expert in solid mechanic) derivations are for an interested reader conveniently summarized in Balek [2].

1. Regularization in nonparametric regression

Let us start by the formulation of the data-analytic problem. Regularization, or "the roughness penalty approach" seeks a nonparametric regression fit, $\hat{f}$, via minimizing

$$L(z, f) + J(f) \looparrowright \min_f ! \tag{1}$$

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, λ can be considered as a part of J, if needed be.

The first part of the minimized function, L, represents 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

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

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

$$J(f) \looparrowright \min_f ! \ \text{ subject to } z_i = f(u_i). \tag{2}$$

The second part of the minimized function, J, is traditionally referred to as the *penalty*. It is responsible for the shape of $\hat{f}$, the characteristic form of the minimizing f. Indeed: divide the process of solving (1) into two stages; in the first stage determine the fitted values, $\hat{f}(u_i)$; the second stage then amounts to solving (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 intending to make the resulting fit smooth is designed to control some derivative(s) of its argument; in the univariate case, the typical expression is

$$J(f) = \int_\Omega \big(f^{(k)}(x)\big)^2 \, dx = \int_\Omega \big(f^{(k)}\big)^2 \, dx \tag{3}$$

—hereafter, we will omit functional arguments in the penalty expressions as indicated by the rightmost part of (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 of square to, say, absolute value may be dictated by reasons of computational convenience. The presence of an integral signifies that the penalty expresses a summary behavior of the derivative. For now, we postpone the discussion of the aspects of the domain of integration Ω.

It may be not entirely clear which order of derivative, k, is to be chosen, and there are in fact almost no guidelines to this end. One of the existing 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*

$$J(f) = \int_\Omega \left(f''\right)^2 dx. \tag{4}$$

This model envisions the graph of f on Ω 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 resumes 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

$$J(f) = \int_\Omega |f''|\, dx = \bigvee_\Omega f', \tag{5}$$

the total variation of f' on Ω. This is the penalty used by Koenker, Ng and Portnoy [19] in their proposal for "quantile smoothing splines" and further cultivated by He, Ng and Portnoy [14] (the equality of the integral to the total variation is the Vitali theorem, in its classical version restricted to absolutely continuous functions; full generality an be achieved by viewing derivatives in the sense of Schwartz distributions). The penalty (5) is a member, for $k = 2$, of the family analogous to (3),

$$J(f) = \int_\Omega |f^{(k)}|\, dx = \bigvee_\Omega f^{(k-1)}, \tag{6}$$

considered by Mammen and van de Geer [22], Koenker and Mizera [17, 18], and others.

The bivariate case brings more possibilities, but the overall picture is similar: elastic deformation energy involves square, or squared norm; plastic energy absolute value, or norm. While the elastic case is well known, the connection of total variation penalties to deformation energy in plasticity may not be that obvious and the relevant variational principles seem not to be available in the existing physical or engineering literature on plasticity theory. As the latter are to an extent analogous to those in elasticity, we start by the review of the latter.

3. Elastic deformation

The notion central in the calculation of deformation energy is *stress*. It is defined as the distribution of the internal tension forces caused by the external forces; for instance, the simplest case of a beam fastened at one end and subject to the action of an aligned load at the other end exhibits (a scalar) stress, σ, which is the ratio of this load and the area of the cross section of the beam (Figure 1).

The relationship of stress to deformation is called *constitutive equation*. It can be used to obtain the expression of the *deformation energy*, $J(f)$, the latter being defined as the work done by the stress in the course of the deformation; its local characteristic, *(deformation) energy density* (the increment of the deformation energy per infinitesimal volume), can be thus obtained via integration involving locally described stress and deformation.

In the beam example above, the deformation is characterized by the relative dilation, Δ, the ratio of the length increase of the extended beam to the length of the original beam. The energy density is then the integral of σ over all values δ of the relative dilation ranging from 0 to Δ. If the constitutive equation is the Hooke law, $\delta = \sigma/E$, then $\sigma = E\delta$, and the energy density is $E\Delta^2/2$. (The constant E is called Young's modulus of elasticity.) The deformation energy is then obtained by integrating this density over the whole body—that is, here by multiplying by its volume.

The applicability of the Hooke law in the beam example is restricted to the situations when load does not exceed certain critical value—we have defined elastic deformations as the reversible ones. The Hooke law then states that deformation is proportional to stress; the relation between stress and deformation is thus linear—if the stress is small (the principle of small causes yielding linear responses; we disregard here a short transitory region of stresses when the deformation is still reversible but the constitutive equation becomes nonlinear).

4. Smoothing with elastic penalties

The physical model for the univariate nonparametric regression 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

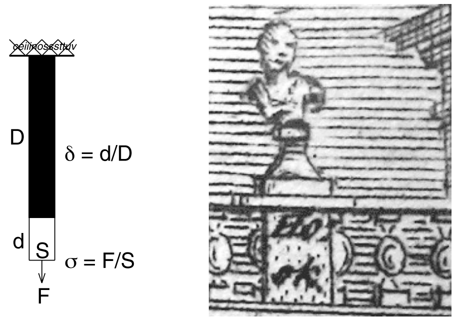

FIG 1. *The beam example (left), illustrating the law of Hooke (right; in fact, no authenticated Hooke's likeness survives, a situation attributed by Arnol'd [1] to certain Newton's efforts).*

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

$$J(f) = \int_\Omega \big(f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2\big)\, dx\, dy, \tag{7}$$

see Green and Silverman [12], Wahba [35]. If ν 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 2ν times the integral of $\det H = h_1 h_2 = f_{xx} f_{yy} - f_{xy}^2$. The deformation energy is then

$$J(f) = \int_\Omega \big(f_{xx}^2 + 2(1-\nu)f_{xy}^2 + f_{yy}^2 + 2\nu f_{xx} f_{yy}\big)\, dx\, dy. \tag{8}$$

An alternative way to arrive to this expression is via rotational invariance arguments, in the spirit of Landau and Lifshitz [21]: 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 an orthogonally invariant function of H; if the energy is presumed to be a quadratic function, such invariants are only $\operatorname{tr} H^2$ 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).

A substance with the Poisson constant ν close to 0 is cork—indeed it exhibits almost no lateral expansion under compression. For typical materials, the values of ν are between 0.3 and 0.4; for instance, steel has 0.30, copper 0.34, gold 0.42. Materials with $\nu < 0$, expanding when stretched, were discovered only recently [20].

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

$$J(f) = \int_\Omega (f_{xx} + f_{yy})^2\, dx\, dy = \int_\Omega (\Delta f)^2\, dx\, dy, \tag{9}$$

which can be traced in the works of O'Sullivan [25], Duchamp and Stuetzle [8], Ramsay [26], and others. It could be also nicknamed "the penalty of Sophie Germain", because it emerged in her first attempt to win the prize of the French Academy of Sciences (in a contest initiated by Napoleon). While the condition $-1 \le \nu \le 1$ follows by the simple requirement of positive definitess 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 ν must not exceed 1/2. Hence the value $\nu = 1$ is "unphysical"[1]. Another distinctive value, $\nu = -1$ yields a penalty

$$J(f) = \int_{\Omega} \big((f_{xx} - f_{yy})^2 + 4f_{xy}^2\big)\, dx\, dy, \tag{10}$$

which seems not yet encountered in a smoothing context. Neither do the instances of (8) with general ν, from the optimization point of view convex combinations of penalties (7), (9), and (10).

To get a better understanding of these penalties, it is instructive to solve the simplest problem in this setting, a circular plate lifted at the center. 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 speaking, this is *not* a problem corresponding to point interpolation (or nonparametric regression)—but can be thought of as an idealization of such, when f interpolates 1 at origin and 0 at many points scattered in some uniform manner along the boundary. Mostly important, however: this problem can often be solved in a closed form, at least partially (and then finished by a numerical method). A first exercise of this kind was worked out for elastic, simply-supported plates by Poisson in 1829.

Consider first the integration domain to be the unit disc with radius $R = 1$. 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

$$J(f) = 2\pi \int_0^1 \left(r w_{rr}^2 + \frac{w_r^2}{r} + 2\nu w_{rr} w_r \right) dr. \tag{11}$$

One can now find w minimizing (11), subject to boundary conditions $w(0) = 1$ and $w(1) = 0$, numerically; for elastic penalties, it is in fact possible to solve the corresponding Euler-Lagrange equation in an analytic form, using standard tricks developed for this type of equations; see Balek [2] for details. The solution is (see Figure 2)

$$\hat{w}(r) = r^2 \left(\frac{2(1+\nu)}{3+\nu} \log r - 1 \right) + 1, \qquad J(\hat{f}) = \frac{8\pi(1+\nu)}{3+\nu}.$$

A possible way to evaluate the results for various ν may derive from a feel that the shape underlying the data is something like the cone centered at the orgin; while we desire the final solution to be smooth, we also do not want it to yield too much to the potential rigidity of the smoothing scheme. In other words, we may want to obtain the smooth shape, but as sharp as possible at 0. This would deter us from $\nu = -1$, interpolating by a quadratic surface, and drive us rather to larger ν, ultimately to $\nu = 1$. However, it has to be said that shapes for various ν differ only very little, so eventually we may well choose ν on the basis of computational convenience, favoring $\nu = 0$.

[1] Sophie's choice is understandable in the light of the fact that the underlying differential equation (corrected by the supportive referee Lagrange) does not involve ν. Despite being the only contestant, she was not awarded the prize on her second attempt, due to 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). Only the third attempt two years later brought her the pledged kilogram of gold.

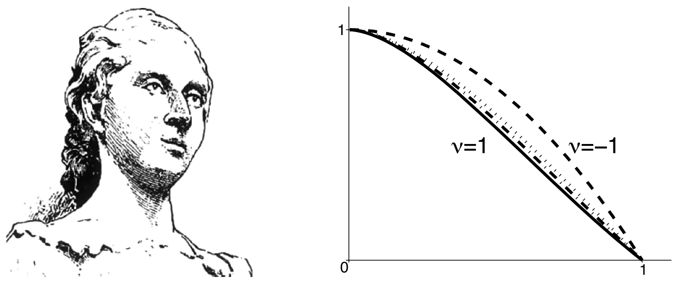

FIG 2. *The sections of solutions (the domain of integration being the disc with radius $R = 1$) for $\nu = -1, 0, 0.42$ and also the value $\nu = 1$, somewhat put forward already by Sophie Germain (left).*

The recognition of the potential of elastic considerations in the interpolation and smoothing of data dates back at least to Sobolev in 1950's. The functional minimized in the variational problem involving a penalty like (7) is quadratic: the Euler-Lagrange equation is thus linear, the solutions have the superposition property and form a linear space—for fixed values of covariates (x_i, y_i) a finite-dimensional one. The structure of this linear space is that of a Hilbert space, which opens a possibility to construct solutions via reproducing kernels. Those were introduced in the smoothing context by Wahba [35], and became popular in the recent statistical machine learning literature—where they opened a whole new range of new possibilities, in which penalties are often secondary to kernels, and the properties of the resulting solutions are inferred rather in an intuitive fashion; see, for instance, Schölkopf and Smola [29].

From the physical point of view, the domain of integration Ω is naturally determined by the solid body represented by f; consequently, it naturally comes as bounded. In the smoothing context, the domain of integration is inessential in the univariate case, where the same results are obtained for any Ω containing all u_i. In the bivariate case, the choice of the domain influences the solution, and thus may become an undesirable detail—unless the smoothing domain comes naturally from the specification of the problem, as for instance in Ramsay [27].

If we enlarge the integration domain in our radial example, for simplicity still retaining it a disc but now with radius $R > 1$, the shape of the plate becomes different. In particular, for $\nu = 0$ we have

$$\hat{w}(r) = \begin{cases} r^2(p \log r - 1) + 1, & \text{when } r \le 1, \\ -p \log r + (p-1)\big(r^2 - 1\big), & \text{when } 1 < r \le R, \end{cases} \tag{12}$$

with $p = (1 + (2R)^{-2})^{-1}$, the optimal value of the penalty equal to $4\pi p$. The resulting shapes can be seen in the left panel of the Figure 3; the tendency with increasing R is similar to that observed with increasing ν.

The necessity to deal with the integration domain was overcome by the idealization idea of Harder and Desmarais [13], perfected by Duchon [9] and Meinguet [24]. They demonstrated that it is possible to take Ω to be the entire $\mathbb{R}^2$, and subsequently obtained expressions for the basic functions in closed form; this substantially facilitates the application of Hilbert-space methods, and reduces the whole problem to a system of linear equations. See also Wahba [35]. The resulting technology became a standard part of the numerical and statistical methodology—the denomination "thin-plate spline" now assumes, as a rule, implicitly $\Omega = \mathbb{R}^2$. Special adjectives are added rather otherwise: for example, Green and Silverman [12]

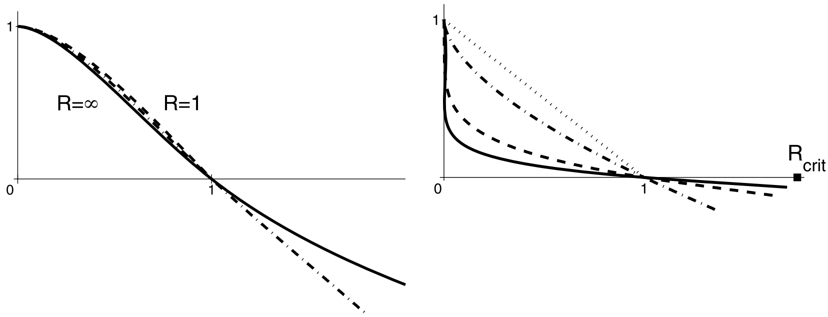

FIG 3. *Elastic and plastic penalties on enlarged domains. Left panel shows the sections of solutions for $\nu = 0$, and $R = 1, 1.8, +\infty$. Right panel shows the sections of the solution of the interpolation problem for the plastic penalty with $\kappa = 0$, on a disc with radius R equal to* $1, 1.375, 1.660, 1.715$.

speak about "finite-window thin-plate splines" in case when Ω is bounded. They motivate their interest in such a variant by the desire to achieve better behavior near the boundary of the data cloud (which may be unduly influenced by adopting the unbounded domain).

In the radial example, the thin-plate spline solution is obtained by letting $R \to +\infty$, making $p \to 1$ in (12). This yields

$$\hat{w}(r) = \begin{cases} r^2(\log r - 1) + 1, & \text{when } r \leq 1, \\ -\log r, & \text{when } r > 1, \end{cases}$$

with the optimal penalty 4π. 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 ν in this case (returning in this case somewhat to the perspective of Sophie Germain).

5. Plastic penalties

As already mentioned, the difference between elasticity and plasticity is reversibility —here understood in its apparent, macroscopic[2] expression: if the body returns to its original configuration, the deformation is elastic; if it does not, then it is (at least partially) plastic.

To identify our position within existing plasticity theories, we need to invoke several notions, pretty much exclusively for the sake of this paragraph. Typically, increasing the forces applied to a body first causes an elastic deformation—unless we face an instance of rather idealized behavior, "rigid plasticity"—until a critical value called *yield limit* is reached, marking the transition to the subsequent plastic behavior. A theory describing the plastic regime in its initial stages, in a manner analogous to elasticity theory, is known as the *deformation theory of plasticity.* It relates stress to the deformation in a way that views plasticity as a kind of nonlinear elasticity. The version of this theory for "perfectly plastic rigid-plastic bodies" provides a physical model for certain alternative penalties in nonparametric

[2]On the microscopic level, the irreversibility originates in the motion of dislocations in the crystallic latice of the body (see Landau and Lifschitz [22]), which is not related to the irreversibility of thermodynamic processes described by Bolzmann's H-theorem, but is rather a consequence of the fact that the energy of the material contains many local minima close to each other; in any case, this aspect is irrelevant for what we are pursuing here.

regression. "Perfect plasticity" refers here to an idealized behavior characterized by the absence of so-called hardening, the necessity to increase stress to produce further plastic deformations. The deformation theory can be extended to capture the latter phenomenon, but its principal shortcoming is its lack of explanation for irreversibility—this motivated other theories of plasticity, in particular incremental, or flow plasticity theory; see Kachanov [16]. All simplifications notwithstanding, the deformation theory of perfect rigid plasticity is extensively employed in engineering, as the basis of so-called limit analysis—the study of limit loads when plasticity takes over (and the subsequent deformations). Simply put, the deformation theory is capable to tell us when structures like roofs and bridges are prone to collapse.

As an illustration, consider again the beam with aligned load. If we increase the stress, the beam extends according to the law $\delta = \sigma/E$; however, after the stress reaches the yield limit $\mathcal{E}$, the beam extends further with the stress staying unchanged. Then, after the beam has been stretched long enough, we need to increase the stress again in order that the dilation increases. The three stages described here are respectively elastic deformation, yielding and hardening. "Rigid plastic" are the idealized materials that skip the first stage; "perfectly plastic" are those that skip the third one. Note that in addition to the constitutive equation $\delta = \sigma/E$, for a perfectly plastic beam we have also the constitutive inequality $\sigma \leq \mathcal{E}$, which must be satisfied when the beam is in equilibrium.

For a rigid-plastic beam, the calculation of deformation energy is even simpler than for an elastic beam. By integrating the stress $\mathcal{E}$ over the values of δ ranging from 0 to Δ we find that the deformation energy is $\mathcal{E}\Delta$. If the load acts in the opposite direction, so that the beam is compressed and the relative dilation is negative, the energy density is $-\mathcal{E}\Delta$. Thus, the general expression for the energy density is $\mathcal{E}|\Delta|$. Consider now a rigid-plastic rod that is neither entirely stretched nor compressed, but bent. Again, as in an elastic rod, the layers on one side of the central surface are stretched and the layers on the other side of the central surface are compressed, so that the mean dilation is zero; however, the mean *magnitude* of the dilation is nonzero, and from the geometry of the problem it follows that it is proportional to the magnitude of the curvature of the beam, or $|f''|$ for small deflections. In this way we find that the deformation energy of the rod is given by (5).

The principal concept we will work with from now on, the cornerstone of the deformation theory, is the *yield criterion*. In two dimensions, it is represented by an orthogonal-similarity-invariant norm $\|\cdot\|$ on 2×2 symmetric matrices. The fact that any such norm can be fully characterized in terms of eigenvalues leads to its two-dimensional representation via the so-called *yield surface*—which visualizes (Figure 4), in the space of eigenvalues, the *constitutive inequality*

$$\|M\| \leq \mu; \tag{13}$$

here M is the moment matrix (related to stress) and μ is a threshold value: the plate is deformed elastically or remains flat if $\|M\| < \mu$, and may be deformed plastically if $\|M\| = \mu$. The *constitutive equation* is subsequently obtained from the postulate that the Hessian, H, is proportional (with negative proportionality constant) to the gradient of $\|M\|$. Consistently with the general concept of duality in the mechanics of deformable bodies, discussed in Témam [32], the ensuing plastic deformation energy corresponding to (13) is

$$\int_\Omega \|H\|_* \, dx \, dy, \tag{14}$$

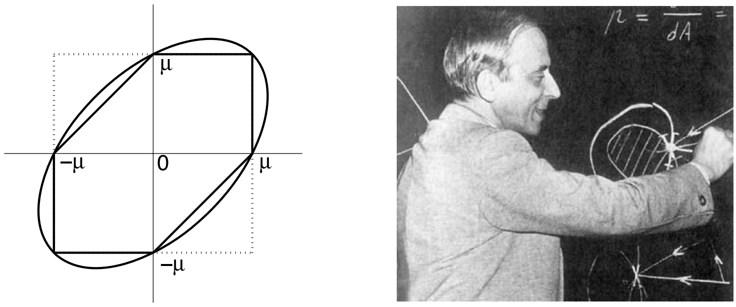

FIG 4. *Various yield criteria. Left: the Tresca hexagon can be seen inscribed in the von Mises (right) ellipse; the dotted line represents the square yield criterion.*

where $\|\cdot\|_*$ is the norm dual (conjugate) to $\|\cdot\|$; see Collatz [5]. The expression (14) can be obtained in a similar way as (8) for the elastic deformation energy, introducing a yield surface in the plane of principal stresses rather than in the plane of principal moments, and formulating the constitutive equation in terms of relative dilations rather than in terms of Hessian matrix. We omit the details, only point out that (14) for a plate is a natural generalization of (5) for a rod.

The yield criterion (the norm and the yield limit μ) characterize plastic properties of a particular material. This begets—in contrast to the elasticity theory, where the material is described by a single number, the Poisson ration ν—much more variety, as witnessed in the engineering literature by the abundance of yield criteria targeted to capture specific virtues of various substances. One of the first, dating back to 1868, is due to Henri Tresca (a mechanical engineer involved in the design of the modern international meter standard). Richard von Mises, following an abandoned 1863 idea of Maxwell, proposed the elliptic criterion that carries his name (sometimes also that of Maximilian Huber, who published some preliminary ideas in his article in Polish). While the original intention was to simplify the analysis, it turned out that the new criterion performed even better when confronted with experimental data.

Another example of a yield criterion, with a particularly simple mathematical expression, is the *square criterion*, deemed good for concrete plates; see Mansfield [23]. It corresponds to the ℓ^∞ norm and its dual to the ℓ^1 norm (see Figure 4). In our context, an interesting family of yield criteria is that proposed by Yang [37], with an objective to achieve better description of materials like marble and sandstone. It is parametrized by a dimensionless parameter $\kappa \in [0, 1/2]$ (a "plastic Poisson constant") and gives similar expressions for deformation energy as in the elastic case,

$$J(f) = \int_\Omega \sqrt{f_{xx}^2 + 2(1-\kappa) f_{xy}^2 + f_{yy}^2 + 2\kappa f_{xx} f_{yy}}\, dx\, dy. \tag{15}$$

The family contains the von Mises yield criterion for $\kappa = 1/2$; from nonparametric regression perspective, it can inspire a family of "plastic" penalties, using not only "physical" κ, but all those satisfying $-1 \le \kappa \le 1$, the requirement ensuring the convexity of the penalty. See Figure 6.

An inspiration in the opposite direction could be the case of the penalty proposed for measuring bivariate roughness by the second author. The penalty amounts to the expected univariate roughness along a randomly oriented segment: a needle of unit length is thrown on the (x, y)-surface; a univariate roughness penalty of $f(x, y)$

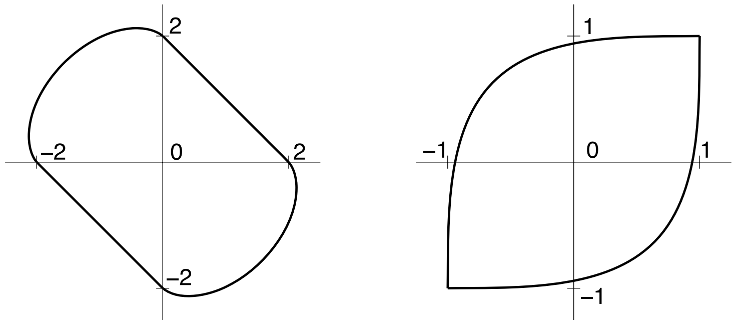

Fig 5. *"Buffon penalty" norm (left), and the associated yield criterion (right).*

evaluated for (x, y) along the needle; and the average of many such evaluations taken. For obvious reasons, this measure of roughness was dubbed "Buffon penalty". The associated matrix norm in the expression (14) bears some similarity with a spectral norm: if the matrix is interpreted as a quadratic form, then the spectral norm returns its maximal value over vectors with unit norm, while the "Buffon norm" returns the average value (with respect to the uniform distribution over the unit sphere of directions). The corresponding norm and its dual are visualized in Figure 5; the yield criterion appears to be a sort of merge between those of von Mises and Tresca.

For plastic penalties with general norms, the solutions of the radial example, the circular plate lifted at the center, are again radial functions and can be expressed in terms of polar coordinates. For instance, for the family (15) we have

$$J(f) = 2\pi \int_0^R \left(r^2 w_{rr}^2 + w_r^2 + 2\kappa r w_{rr} w_r\right)^{1/2} dr. \tag{16}$$

However, the optimal $\hat{w}$ have to be generally computed numerically, only certain special cases can be solved in closed form. The integration domains are again discs with radius R. The easiest case is $R = 1$ and $\kappa = 0$, when $w(r) = 1 - r$, for r going from 0 to 1. In fact, this is a special case: for any $R \in [1, R_{\text{crit}})$ and any κ, there is a unique $q \in [0, 1)$ such that

$$w(r) = 1 - r^{1-q} \qquad \text{for } 0 \leq r \leq 1,$$

and these solutions can be extended to r ranging from 1 to R; we omit the details. Unfortunately, the case $R = +\infty$, the extension of the integration domain Ω to the whole $\mathbb{R}^2$, does not work out here as in the elastic case: starting with some critical value R_{crit} (which appears to be 1.765) of R (that is, *not* asymptotically), the solutions degenerate: they are equal to 1 at 0 and to 0 elsewhere. The right panel of Figure 3 clearly indicates this tendency, showing the sections with $q = 0.4, 0.8, 0.9$, corresponding approximately to $R = 1.375, 1.660, 1.715$.

Albeit the introduction of total variation penalties (6) by Koenker, Ng and Portnoy [19] in nonparametric regression, and by Rudin, Osher and Fatemi [28] in image reconstruction predated the recent interest in regularization with ℓ^1 penalties, it may be appealing to compare the latter, made popular by Chen, Donoho and Saunders [4] and Tibshirani [33], to the plastic penalties discussed here. Certain plain explanations (and new labels like "fused lasso" for total-variation regularization [11]) are particularly inviting in the univariate case: the piecewise-linear form of the solutions for the penalty (5) may be interpreted as a consequence of the tendency of the absolute value to pick up “sparse” solutions for f''. It is only the bivariate case where the differences stand more out (although even there is a possibility to interpret the plastic penalties in the “group lasso” vein). On the other hand, there is some analogy to sparsity aspects even there, demonstrating itself in the nonparametric regression context as the possibility of capturing qualitative features like spikes and edges, without oversmoothing the relevant information; see Davies and Kovac [6] for an overview of this aspect.

In the ℓ^1 context, however, the relevant linear spaces of solutions are no longer Hilbert spaces; the possibility of using reproducing kernels is lost, and with it possibly one potential attraction of unbounded integration domains. Even in the ℓ^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 reproducing kernels are non-sparse—and there does not seem to be any workaround for this; see Silverman [30], 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 [15], Duchamp and Stuetzle [8], Ramsay [26], and others. The algorithms thus return back to that of boundary value problems, where the idealization of Ω to $\mathbb{R}^2$ is not any longer a necessity.

6. Extensions and ramifications

In the final section, we would like to touch briefly some other aspects where physical analogies may or might provide some guidance. The first of them concerns blends—convex combinations of several penalties. In the discussion of the results of smoothing with elastic penalties for the radial example we noted that, from a certain perspective, it may be desirable to have those fits as “sharp” as possible at 0. As it was felt that purely elastic penalties may not sufficiently achieve this objective, several solutions were proposed to this end.

One possibility is to use the plastic penalty, which in the radial example with $R = 1$ yields the “ideal” conic shape of the fit. However, its dependence on the integration domain, impossibility of an idealized extension to the unbounded domain, and in particular, the lack of Hilbert-space methods for finding solutions, motivated so-called *thin-plate splines with tension.* Recall that in our mechanic deliberations, the thin plate was supposed only to bend, not to stretch. In a more realistic situation involving some stretching and some bending, the physical model results in a “blended” penalty: for instance, a convex combination of (7) and the Dirichlet penalty

$$J(f) = \int_\Omega \left(f_x^2 + f_y^2\right) dx\, dy. \tag{17}$$

The latter coresponds in mechanics not to an 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. It is known that it is not possible to use (17) in smoothing by itself—an instance of a broader phenomenon of increasing need for derivatives in regularization penalties, related in some sense to the embedding theory of functional spaces of Sobolev type, an intuitive explanation of which is given in terms of the radial interpolation example on page 160 of Green and Silverman [12].

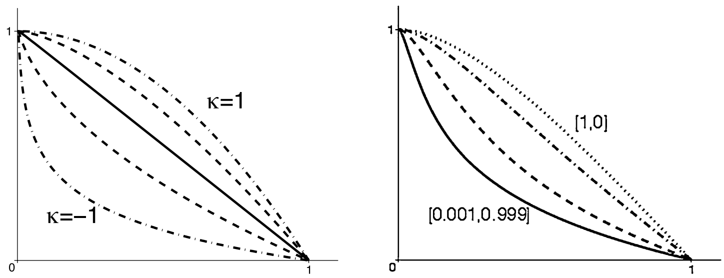

FIG 6. *Plastic splines vs. thin-plate splines with tension. Left: plastic penalties* (15)*, with* $\kappa = 1, 1/2, 0, -1/2, -1$*. Right: convex combinations of thin-plate spline penalty* (7) *with tension term weighted* $0, 0.9, 0.99, 0.999$*.*

In the radial example minimizing (17), the solution degenerates—more precisely, it does not exist. However, as can be seen in the right panel of Figure 6, adding even a small portion of an elastic penalty acts as a regularizer, and a fair proportion of "tension" yields more pronounced peak in the solution of the radial interpolation (an excessive one draws it closer to the funnel-like degeneracy). Moreover, the solutions in Figure 6 can be actually obtained in closed form using modified Bessel functions, via (11) and the analogous expressions for (17); *thin-plate splines with tension* proposed by Franke [10] actually provide quite accurate approximation to this solution, especially for higher tension weights.

A possible suggestion in a similar direction is to regularize plastic by an elastic penalty. A way that would be analogous to the approach of thin-plate splines with tension is to consider a convex combination of a plastic and elastic penalty. The results can be seen in the left panel of Figure 7; it turns out that the regularizing effect (and thus a possibility of extending the whole formulation to the unbounded domain) is achieved by a quite small proportion of the elastic penalty (in fact, some optimization algorithms may already exhibit a similar regularization effect spontaneously, due to the nature of the approximations involved). It may be of interest that a similar idea of combining ℓ^1 and ℓ^2 penalty appeared in recently in a different context under the name "elastic net" [38].

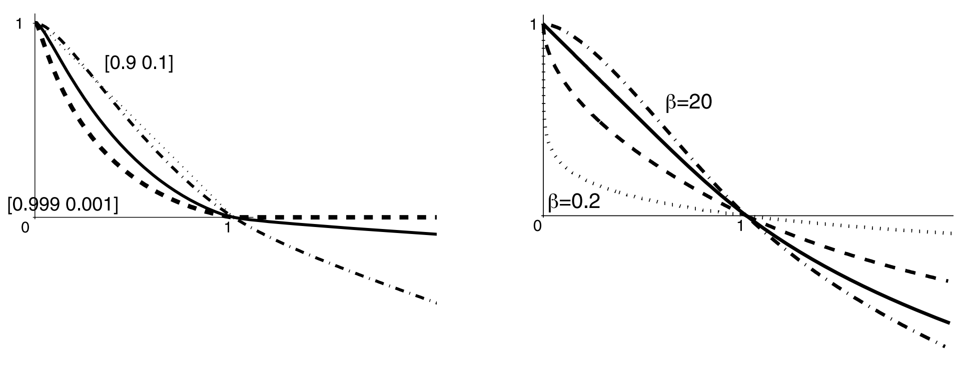

FIG 7. *Elastoplastic penalties. Left: total-variation penalty* (6) *regularized via the convex combination with thin-plate spline penalty* (7) *weighted* $0.1, 0.01, 0.001$*, on the disc with radius* $R = 2.3$*. Right: elastoplastic penalty with* $\iota = 0$ *and* $\beta = 20, 2.845, 1, 0.2$*. The dotted line in the left panel is the result for the total-variation penalty when* $R = 1$*, which in the right panel almost coincides with the fit for* $\beta = 2.845$*.*

It would be tempting to call the blended penalty "elastoplastic"; however, if a situation—corresponding even more to reality—when there is some plastic and some elastic behavior in bending, is approached from the physical point of view, the object we arrive to is Hencky's plate, the perfectly plastic elastoplastic plate described by the deformation theory. In such bodies, a pure elastic deformation takes place if the stresses lay inside the yield diagram, and a combined elastic and plastic deformation possibly takes place if the stresses are placed at the surface of the yield diagram. Furthermore, the plastic part of deformation is the same as in rigid-plastic bodies. The leads to a family of penalties

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

generalizing, for $\iota = \nu = \kappa$, the elastic penalties (8), with $\varphi(u) = u^2$, and plastic penalties (15), with $\varphi(u) = |u|$, via a function φ known in statistics as the Huber function, equal to $u^2/2$ for $|u| \leq \beta$, and to $\beta|u| - \beta^2/2$ otherwise; the parameter β is determined by the yield criterion. The results for the radial example, for $\iota = 0$ and various choices of β, can be seen in the right panel of Figure 7.

References

[1] Arnol'd, V. I. (1990). *Huygens and Barrow, Newton and Hooke.* Birkhäuser, Basel.

[2] Balek, V. (2011). Mechanics of plates. http://xxx.lanl.gov/abs/1110.2050.

[3] Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press, Cambridge.

[4] Chen, S. S., Donoho, D. L. and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61.

[5] Collatz, L. (1966). *Functional Analysis and Numerical Mathematics.* Academic Press, New York.

[6] Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Ann. Statist.* **29** 1–65.

[7] Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis.* Wiley, Chichester.

[8] Duchamp, T. and Stuetzle, W. (2003). Spline smoothing on surfaces. *Journal of Computational and Graphical Statistics* **12** 354–381.

[9] Duchon, J. (1976). Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *R.A.I.R.O., Analyse numérique* **10** 1–13.

[10] Franke, R. (1985). Thin plate splines with tension. *Computer Aided Geometric Design* **2** 85–95.

[11] Friedman, J., Hastie, T., Höfling, H. and Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Statist.* **1** 302–332.

[12] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London.

[13] Harder, R. L. and Desmarais, R. N. (1972). Interpolation using surface splines. *J. Aircraft* **9** 189–191.

[14] He, X., Ng, P. and Portnoy, S. (1998). Bivariate quantile smoothing splines. *J. R. Statist. Soc. B* **60** 537–550.

[15] Hegland, M., Roberts, S. and Altas, I. (1998). Finite element thin plate splines for surface fitting. In: *Computational Techniques and Applications:*

CTAC 97 (J. Noye, M. Teubner AND A. Gill, eds.). World Scientific, Singapore.

[16] KACHANOV, A. M. (1956). *Osnovy Teorii Plastichnosti.* GITTL, Moscow.

[17] KOENKER, R. AND MIZERA, I. (2002). Elastic and plastic splines: Some experimental comparisons. In *Statistical Data Analysis Based on the L1-Norm and Related Methods* (Y. Dodge, ed.) 405–414. Birkhäuser, Basel.

[18] KOENKER, R. AND MIZERA, I. (2004). Penalized triograms: Total variation regularization for bivariate smoothing. *J. R. Statist Soc. B* **66** 1681–1736.

[19] KOENKER, R., NG, P. AND PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680.

[20] LAKES, R. (1987). Foam structures with a negative Poisson's ratio. *Science* **27** 1038–1040.

[21] LANDAU, L. D. AND LIFSHITZ, Y. M. (1986). *Theory of Elasticity.* Pergamon Press, London. [Third English edition, revised and enlarged by E. M. Lifshitz, A. M. Kosevich and L. P. Pitaevskii; translated from Teoriia uprugosti, Moskva, Nauka, 1986, by J. B. Sykes and W. H. Reid].

[22] MAMMEN, E. AND VAN DE GEER, S. A. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413.

[23] MANSFIELD, E. H. (1957). Studies in collapse analysis of rigid-plastic plates with a square yield diagram. *Proc. R. Soc. of London A* **241** 311–338.

[24] MEINGUET, J. (1979). Multivariate interpolation at arbitrary points made simple. *J. Appl. Math. Phys. (ZAMP)* **30** 292–304.

[25] O'SULLIVAN, F. (1991). Discretized Laplacian smoothing by Fourier methods. *J. Amer. Statist. Assoc.* **86** 634–642.

[26] RAMSAY, T. O. (1999). *A Bivariate Finite-Element Smoothing Spline Applied to Image Registration.* Queen's University, Kingston. Thesis.

[27] RAMSAY, T. (2002). Spline smoothing over difficult regions. *J. R. Statist. Soc. B* **64** 307–319.

[28] RUDIN, L. I., OSHER, S. AND FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D* **60** 259–268.

[29] SCHÖLKOPF, B. AND SMOLA, A. J. (2002). *Learning With Kernels.* MIT Press, Cambridge, MA.

[30] SILVERMAN, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc. B* **47** 1–52.

[31] SMALL, C. G. (1996). *The Statistical Theory of Shape.* Springer-Verlag, New York.

[32] TÉMAM, R. (1983). *Problèmes Mathématiques en Plasticité.* Bordas, Paris. [English translation: Mathematical Problems in Plasticity, Bordas, Paris, 1985].

[33] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B* **58** 267–288.

[34] TIMOSHENKO, S. P. AND WOINOWSKY-KRIEGER, S. (1959). *Theory of Plates and Shells.* McGraw-Hill, New York.

[35] WAHBA, G. (1990). *Spline models for observational data. CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, Philadelphia.

[36] WOOD, S. N. (2003). Thin-plate regression splines. *J. R. Statist. Soc. B* **65** 95–114.

[37] YANG, W. H. (1980). A useful theorem for constructing convex yield functions. *J. Appl. Mech.* **47** 301–303.

[38] ZOU, H. AND HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B* **67** 301–320.