Elastic and plastic splines: some experimental comparisons

Roger Koenker and Ivan Mizera

Abstract. We give some empirical comparisons between two nonparametric regression methods based on regularization: the elastic or thin-plate splines, and plastic splines based on total variation penalties.

1. Nonparametric regression with elastic and plastic splines

We seek f to fit the dependence of response points $z_i \in \mathbb{R}$ on two-dimensional covariates (x_i, y_i) living in the domain Ω. To this end, we employ regularization: f is obtained as a minimizer of

$$\sum_{i=1}^{n} \rho(z_i - f(x_i, y_i)) + \lambda J(f). \tag{1}$$

The first term is traditionally called *(in)fidelity*, since it measures the overall lack of fit of $f(x_i, y_i)$ to z_i. The second *penalty* term shrinks the solution towards a more plausible or desirable alternative, the extent of this shrinkage being controlled by the *regularization parameter* λ. See Green and Silverman (1994), Eubank (1999) or Wahba (1990).

1.1. Elastic splines

We coin the term *elastic splines* for what are usually called thin-plate splines (defined on the idealized domain $\Omega = \mathbb{R}^2$). They arise as solutions of (1) with the penalty

$$J_2(f) = \iint_{\mathbb{R}^2} f_{xx}^2(x, y) + 2f_{xy}^2(x, y) + f_{yy}^2(x, y)\, dx\, dy. \tag{2}$$

This penalty can be considered a natural extension of the easily interpretable one-dimensional prototype $\int (f'')^2$. The only unnatural feature is the fact that the penalty is evaluated over all of $\mathbb{R}^2$ instead of over a more realistic bounded domain Ω, for instance the convex hull of the (x_i, y_i) points. The latter alternative was considered, among others, by Green and Silverman (1994) who coined the name *finite-window thin-plate splines*. However, the algorithm for the finite-window alternative is more involved, though not necessarily slower—and, mostly important, not available to us at the present moment, therefore this version of elastic splines will not be considered further here.

The name "elastic splines" comes from the quite well-known physical model underlying the whole setting, in which the penalty is interpreted as the potential energy of a displacement, from the horizontal position, of an elastic thin (metal) plate, the displacement that mimicks the form of the fitted function interpolating the data points. As usual in physical theories, idealizations are inevitable. The displacement should be small, we may say infinitesimal, thus rather in the form $\epsilon f(x,y)$ than $f(x,y)$. The material aspects of the plate are rather limiting with respect to its physical reality: it is thin, that is, we may abstract from its third dimension; it is elastic, hence it does not deform, only bend, and it is a plate, not a membrane, which means it is stiff—its behavior is rather that of steel than that of gum.

Despite its simplifications, such a physical analogy serves as a very useful hint in the world of otherwise potentially endless possibilities.

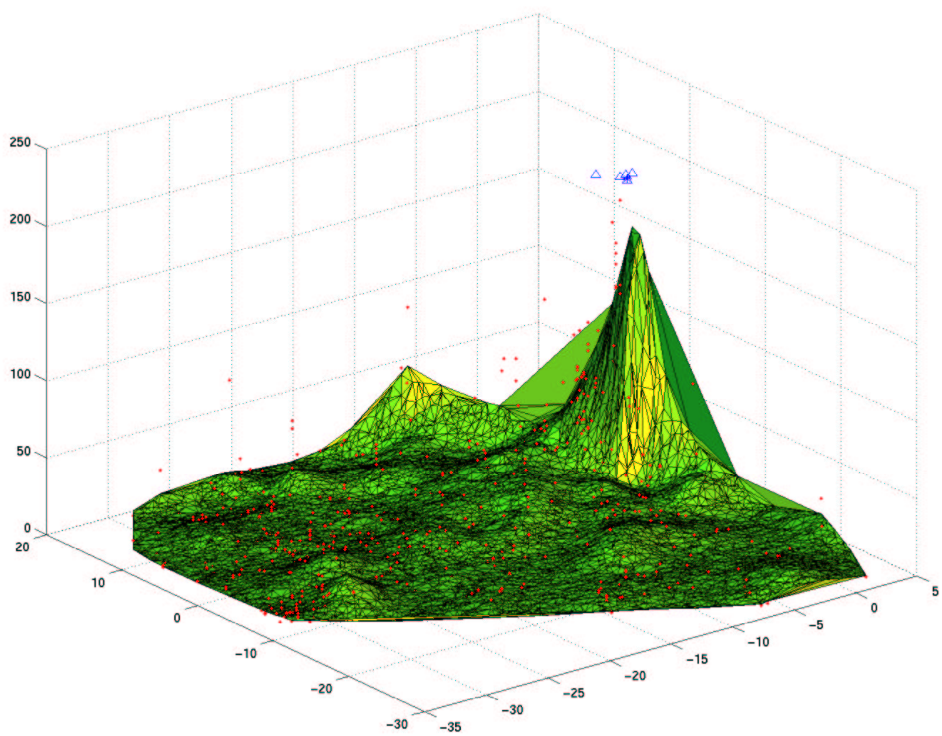


FIGURE 1. Chicago data: elastic fit. The large triangles at rear are artifacts of the visualization method.

1.2. Plastic splines

Plastic splines arise as solutions of another instance of the regularization scheme (1), when the penalty is

$$J^2_{2,\|\cdot\|}(f,\Omega) = \bigvee_{\Omega} \nabla^2 f = \iint_{\Omega} \|\nabla^2 f(x,y)\|\, dx\, dy, \tag{3}$$

where $\nabla^2 f(x,y)$ denotes the Hessian of f at (x,y). Formulas like this should be read with some caution here: the derivatives are not only the classical ones applied to classical smooth functions, but also the generalized ones applying to certain Schwartzian distributions. Probably the easiest way to apprehend this is to think about the right-hand side of (3) as about a definition for smooth functions, which is subsequently extended, by continuity or rather lower semicontinuity, to all functions whose gradient has bounded variation (the property essentially equivalent to the bounded area of the graphs of the components). The similar extension exercise with the quadratic penalty (2) would not yield anything new, but here it considerably broadens the scope and adds also functions with sharp edges and spikes.

The necessity of choosing a matrix norm means that it is more appropriate to refer to (3) as to a family of penalties. Plastic penalties are always considered over bounded Ω; an obvious choice of the matrix norm is the ℓ_2 (Hilbert, Schmidt, Frobenius, Schur) norm $\|\cdot\|_2$. It establishes a parallel between thin-plate and plastic penalties—compare (2) with

$$J^2_{2,2}(f,\Omega) = \iint_{\Omega} \sqrt{f^2_{xx}(x,y) + 2f^2_{xy}(x,y) + f^2_{yy}(x,y)}\, dx\, dy. \tag{4}$$

Despite its appealing simplicity, there are also other, and not irrelevant, norm choices possible. We require that all norms are orthogonal-similarity invariant, to ensure that the resulting penalties are coordinate-free, as is the thin-plate penalty (2). For further motivation, theory and examples, see Koenker and Mizera (2001) and Mizera (2002).

Any plastic penalty, regardless of the choice of the norm, can be considered a natural extension of the one-dimensional penalty equal to the total variation of the derivative, that is, $\int |f''|$ for smooth functions. The latter penalty was introduced by Koenker, Ng and Portnoy (1994), who were motivated by the quantile regression infidelity $\rho(u) = \rho_\tau(u) = u(\tau - I[u < 0])$. Another way to justify the ℓ_1 is on computational grounds, or by scale equivariance considerations as those given by Koenker and Mizera (2001).

Given all of this, and also the form of the plastic penalty, it can be said that plastic splines are an ℓ_1 alternative to the ℓ_2 elastic ones. The adjective "plastic" comes here from the fact that, again under certain idealization, the penalty can be interpreted as the deformation energy, the work done by the stress in the course of deformation, of the plate displacement mimicking the interpolated shape. The theory in which is this interpretation possible, the deformation theory of a perfectly plastic rigid-plastic body, is a kind of a limit case of various other physical approaches to plasticity—nevertheless, it is extensively used in engineering despite all its simplifications. Various yield criteria (a notion considered by the theory) then lead to various matrix norms for plastic penalties. We may again think of a real-world material model such as a metal, but now formable—cast into artistic shapes in the copper foundry.

Another link, connected to the mathematical expression of the penalty, leads to the so-called *total-variation based denoising* of Rudin, Osher and Fatemi (1992), motivated by a desire to recover edges, extrema, and other "sharp" features, while not penalizing smoothness. For parallels in statistics, see Davies and Kovac (2001).

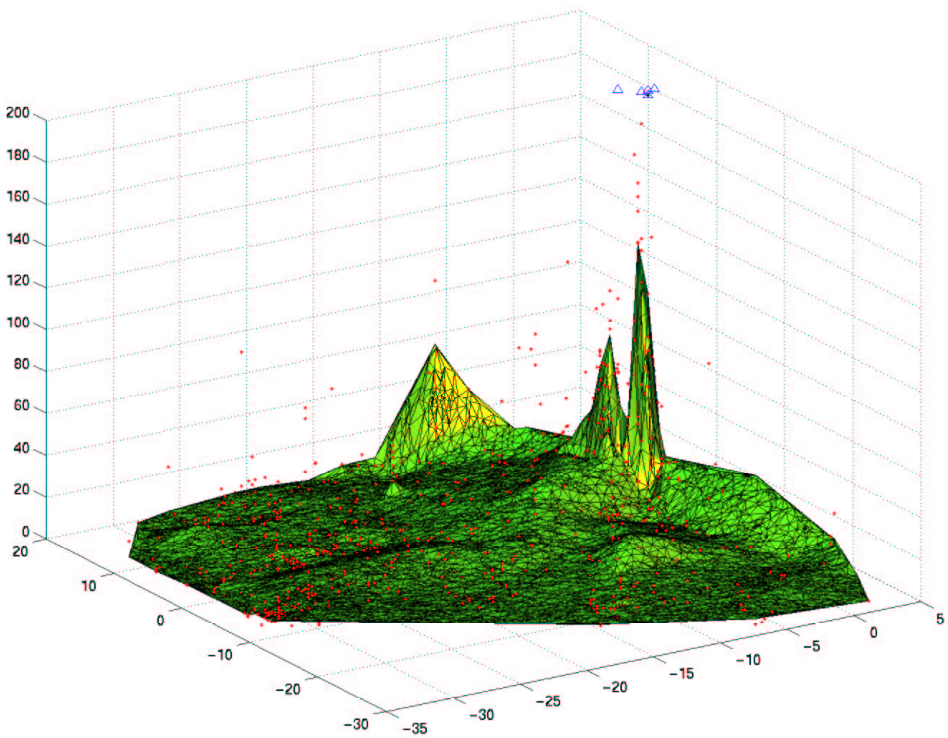


Figure 2. Chicago data: fit by a plastic, triogram spline.

1.3. Objective

The objective of this paper is to give some numerical comparisons between elastic and plastic approach to regularization. "Numerical" stems from the fact that the underlying theory is under development. As mentioned above, the elastic splines are represented by the idealized domain thin-plate splines computed by our MATLAB implementation of the standard algorithm described in Wahba (1990), Gu, Bates, Chen and Wahba (1989), and Green and Silverman (1994). (We have some reasons to believe now that the finite-window versions would even amplify some features of the idealized ones.)

Plastic splines are computed via penalized triogram algorithm, which can be interpreted as the Lagrange finite-element method. The solutions are approximated by functions piecewise-linear on a triangular tesselation of the domain Ω, whose vertices encompass all covariate points, but also many additional *dummy vertices* whose fitted response does not contribute to the infidelity, but their presence increases flexibility of the interpolated surface. All plastic penalties applied to triograms yield the same result (modulo multiplication by a constant which is inessential here):

$$\sum_{e \in E} \|\nabla_1(e) - \nabla_2(e)\| \ \mathrm{length}(e),$$

where E is the collection of all edges dividing two triangles and $\nabla_1(e)$, $\nabla_2(e)$ stand for the gradients in the triangles adjacent to the edge e. This form enables us to formulate the minimization as a linear programming problem, benefiting from the fast implementation of Koenker and Portnoy (1997). For more motivations and the complete description of the triogram algorithm see Koenker and Mizera (2001).

2. The Tipi as an interpolation exercise

It is well-known that interpolation is of crucial importance for determining the form of the general regularization solutions—any solution of the regularization problem is a minimum penalty interpolant of its fitted points. In this section we will briefly describe a canonical interpolation problem, that of interpolating a tipi. Interpolating a *tipi* means finding a function interpolating several points with the z coordinate 0 lying on the unit circumference and one point lying at the origin with value 1.

2.1. The Poisson tipi

The version of the tipi admitting a rigorous solution is an idealized one: the value 1 at the origin remains, while the value 0 is now prescribed on the whole unit circumference. Strictly speaking, it is not a problem falling under the scheme (1), but we may closely approximate it when we set in the non-idealized version, say, 100 equispaced points on the unit circumference equal to 0. The idealized tipi is the cornerstone example in the theory of elasticity, where its form was for the first time derived by Poisson in 1829.

The idealization of the tipi makes it a boundary value problem, which is by rotational invariance solvable by separation of variables. The details can be found in Mizera (2002); here we present just the visualization in Figure 3. The results for the finite-window and idealized thin-plate setting are at its left-hand side. The other three are solutions for plastic penalties with various norms, obtained as a surface of revolution of the curve $g(r) = 1 - r^{1+\kappa}$ with $r \in [0, 1]$. The constant κ is obtained as t minimizing the function $\|\mathrm{diag}\{1, t\}\|$, where $\|\cdot\|$ is the matrix norm

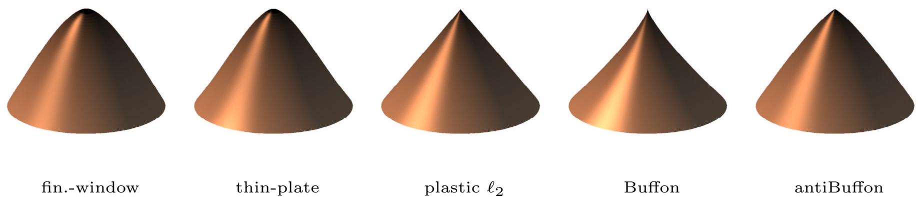


FIGURE 3. Interpolation of the idealized Poisson tipi.

involved in the plastic penalty. The first from left, in the middle of Figure 3, comes the solution for the ℓ_2 matrix norm with $\kappa = 0$; next comes the solution for so-called Buffon variation with $\kappa = -0.1954$ and finally the solution for hypothetical "antiBuffon" variation with $\kappa = 0.1954$.

2.2. A real tipi

For comparison, we interpolated a "real" tipi, with 100 equispaced points on the unit circumference set to 0. The results for thin-plate and triogram algorithms are respectively on left-hand and right-hand side of Figure 4. The results for the elastic case confirm the validity of the thin-plate algorithm; note the intriguing similarity of the triogram interpolant with the "antiBuffon" tipi (there is some heuristic justifications for that).

FIGURE 4. Numerical interpolation of a real tipi. The observed texture is again an artifact of the plotting technique.

3. Green and Silverman data

The data, on the width of the ore-bearing layer in a region of northwest Australia, have been taken from Green and Silverman (1994), where several fits for the idealized and finite-window versions of thin-plate splines are given, for various values of the regularization parameter λ. Here we present, at Figure 5, elastic fits, the idealized thin-plate fit for λ close to 0 (interpolation) and λ selected by generalized cross-validation (smoothing), and related plastic fits; in the latter ones, the

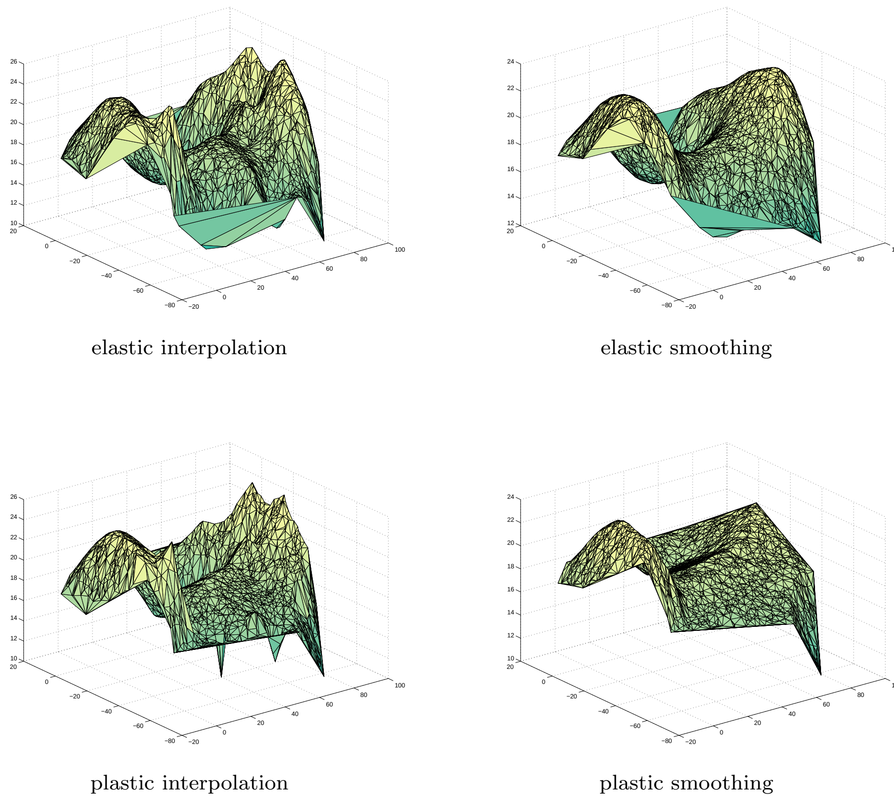


FIGURE 5. Elastic and plastic interpolation and smoothing of the Green and Silverman data.

value of λ was selected to yield approximately the same level of smoothing—the ℓ_1 infidelity roughly equal to the square root of the ℓ_2 one in the elastic fitting.

4. Chicago real estate data

The data, described in Koenker and Mizera (2001), consist of the prices per square foot of 761 sales of undeveloped lots in the Chicago metropolitan area. A piecewise linear fit on the Delaunay triangulation of the original covariate points was given in Koenker and Mizera (2001), the analysis was further refined in Koenker and Mizera (2002). The plastic-spline fit given at Figure 2 involves about 8000 more dummy vertices, which allow us to approximate the plastic spline solution in the finite-element method vein. The thin-plate spline fit is shown at Figure 1; the

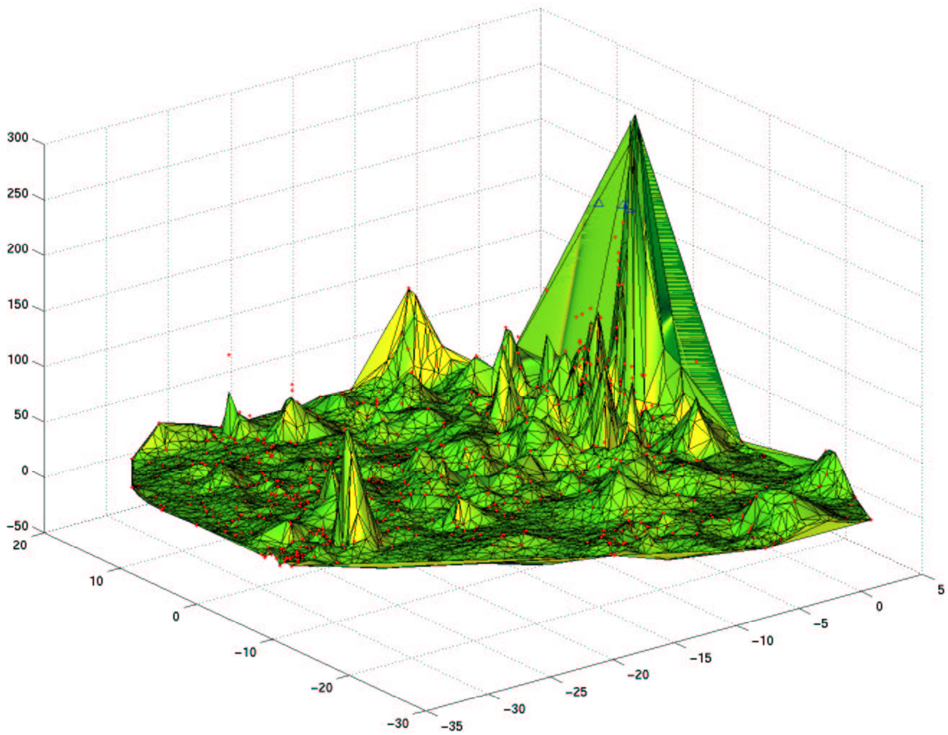
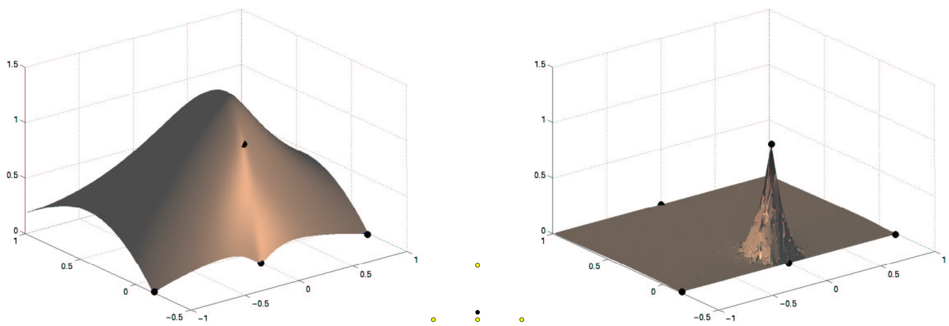


FIGURE 6. Chicago data: elastic splines with λ selected by generalized cross-validation.

regularization parameter λ was tuned rather by hand, after observing the result of generalized cross-validation, shown for completeness at Figure 6. The figures include also the original data points, (visible if not obscured by their surroundings); in particular, the triangles represent the five data points with the largest price, which we had to truncate for plotting purposes, to prevent the inflation of the scale and the subsequent visual triviality of the whole result. Undoubtedly, the figures represent a very incomplete view of the complete outcome; nevertheless, one can draw some feeling (and even more than that if given the possibility of rotating the plots) how the both methods localize the extrema in the data.

5. Another interpolation exercise: an asymmetric peak

Inspired by the Chicago real estate data, we contrived another interpolation exercise: four points, with coordinates $(-0.9877, -0.1564)$, $(0.9877, -0.1564)$, $(0, -0.1564)$ and $(0, 1)$ are set to 0; the point $(0, 0)$ to 1 (see the middle panel of Figure 7). The results of elastic and plastic interpolation are shown at the left-hand and right-hand panel of Figure 7, respectively.



elastic plastic

FIGURE 7. Interpolation of an asymmetric peak.

6. Conclusion

It is premature to draw any serious conclusion. We would like to emphasize that we do not view elastic and plastic splines as competitors. Though both may deal with similar data, each technology has its own specific merits and areas of application. To obtain interpolants for which smoothness is an ultimate desideratum, the quadratic, elastic methods are and probably will remain the best option. On the other hand, when the objective is to isolate local extrema, sharp edges, spikes and similar phenomena in the data, then plastic splines should be considered as a promising alternative.

Acknowledgments

We are grateful to Yadolah Dodge for his patience, and to NSF and National Scientific and Engineering Research Council of Canada for the support.

References

DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–48.

EUBANK, R. L. (1999). *Nonparametric regression and spline smoothing (2nd edition)*. Marcel Dekker, New York.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, London.

Gu, C., Bates, D. M., Chen, Z. and Wahba, G. (1989). The computation of generalized cross-validation functions through Householder tridiagonalization with applications to the fitting of interaction spline models. *SIAM J. Matrix. Anal. Appl.* **10** 457–480.

Koenker, R. and Mizera, I. (2001). Penalized triograms: total variation regularization for bivariate smoothing. Preprint.

Koenker, R. and Mizera, I. (2002). Comment to Spline adaptation in extended linear models by M. H. Hansen and C. Kooperberg. *Statist. Sci.* in press.

Koenker, R. and Portnoy, S. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680.

Mizera, I. (2002). Plastic splines I: data analysis.

Rudin, L. I., Osher, S. and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D* **60** 259–268.

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