

INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

From Data to Assessments and Decisions: Epi-Spline Technology

Johannes O. RoysetRoger J-B Wets

To cite this entry: Johannes O. RoysetRoger J-B Wets. From Data to Assessments and Decisions: Epi-Spline Technology. *In* INFORMS TutORials in Operations Research. Published online: 27 Oct 2014; 27-53.
<https://doi.org/10.1287/educ.2014.0126>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

From Data to Assessments and Decisions: Epi-Spline Technology

Johannes O. Royset

Department of Operations Research, Naval Postgraduate School, Monterey, California 93943,
joroyset@nps.edu

Roger J-B Wets

Department of Mathematics, University of California, Davis, Davis, California 95616,
rjbwets@ucdavis.edu

Abstract Analysts in every field face the challenge of how to best use available data to estimate performance, quantify uncertainty, and predict the future. The data are almost never “just right,” but rather scarce, excessive, corrupted, uncertain, and incomplete. External information derived from experiences, established “laws,” and physical restrictions offer opportunities to remedy the situation and should be utilized. Applications in sustainable energy, natural resources, image reconstruction, financial planning, uncertainty quantification, and reliability engineering are rich with problems where decisions rely on data analysis under such circumstances. We address these problems within a framework that identifies a function that according to some criterion best represents the given data set and satisfies constraints derived from the data as well as external information. Epi-splines provide the linchpin that allows us to handle shape restrictions, information growth, and approximations.

Keywords splines; epi-splines; epi-convergence; epi-topology; financial curves; forecasting; regression; shape restriction; information fusion; constrained optimization

1. Introduction

Much of operations research and management science practice centers on utilizing available data to gain insight about a phenomenon, predict future performance, and instantiate optimization and simulation models. The process intersects several disciplines. Computer scientists refer to it as “learning.” Statisticians talk about “data analysis,” “regression,” and “inference.” At a fundamental level, the process relates to function approximation theory of mathematics, which has a long and prestigious history. Initially, polynomial-type approximations and generalized Fourier series were the mainstay of approximation theory both at the theoretical and computational levels. With the advent of digital computers, piecewise polynomials in the form of splines became computationally tractable and offered the possibility of much more accurate and stable approximations.

Although observed data about a phenomenon receives the primal attention, at least in theoretical studies, additional information derives from a much wider range of sources. Data are never obtained in a vacuum, but rather in a context that provides *external information* in a variety of ways. Physical laws about the phenomenon may dictate that a regression line is increasing. Asymptotic theory could indicate that a probability density function is “nearly” normal. An analyst’s experience gives a host of “soft” information, maybe about ranges of possible values, signs of correlations, and number of modes. It is essential, especially when the data are limited, corrupted, and otherwise deficient, to utilize *all* available information, both from data and external sources. This is intuitively understood and widely exploited in practice. However, the theoretical foundation and computational tools for carrying out this

process are incomplete. In this tutorial, we describe recent efforts to remedy the situation through the formulation of optimization problems and the examination of such problems via variational analysis. Specifically, we address the problem of how to *identify a function that best represents data and also satisfies information-driven constraints*.

The problem arises broadly in curve fitting, interpolation, regression, density estimation, and numerous other contexts. In §2, we describe applications in the financial and commodity markets and list several others. Numerical examples are given in §6. The information-driven constraints are formulated based on external information as well as the data and come in a multitude of forms as illustrated below. The “best” is quantified by a criterion that could be least squares, maximum likelihood, minimum error measure, etc. We refer to this problem as the *function identification problem (FIP)*.

The generality of the FIP and its central position in a wide array of quantitative fields give rise, naturally, to a truly immense literature on the subject. We make no attempt to present a comprehensive treatment, but instead we offer a biased view grown out of our gradual realization of the central role of variational analysis in problems of this kind. We formulate the FIP as a constrained infinite-dimensional optimization problem that represents an *actual* problem. Incomplete information and simplifications introduced for computational reasons lead to *approximate* optimization problems. These problems are subsequently solved by standard, or possibly specialized, optimization algorithms to obtain best fits, estimates, and approximations. We then rely on variational analysis to examine whether the solutions of the approximate problems are indeed approximations of solutions of the actual problem, study associated convergence rates, and obtain other insights. We refer to solutions of an actual problem as *actual functions*.

Since determining an arbitrary function requires pinning down an infinite number of parameters, an unavoidable simplification comes from replacing such functions by simpler ones that are fully characterized by a finite number of parameters. The consideration of simple, but hopefully representative, functions is of course common in linear regression, parametric density estimation, interpolation, and many other areas. *Epi-splines* are newly developed piecewise polynomial functions, described by a finite number of parameters, that are exceptionally flexible. In fact, epi-splines approximate essentially any function one can reasonably expect to encounter in practice, including those with discontinuities, unbounded derivatives, and even the function values $-\infty$ and ∞ . We provide an introduction to epi-splines and their many applications; see Royset and Wets [19, 20, 21] and references therein for further details.

The remainder of the overview is organized as follows. Section 2 provides a glimpse of the many applications served by epi-spline technology. The mathematical formulation of the FIP and a definition of epi-splines are given in §3. The examination of the FIP relies on variational analysis, with pertinent facts reviewed in §4. Section 5 summarizes theoretical results that justify the application of epi-spline technology. The tutorial ends in §6 with numerical examples.

2. Applications

Despite their recent discovery, epi-splines are applied in numerous areas. The first explicit use of epi-splines, under the name epi-curves, was in the context of deriving the zero curves associated with a family of financial instruments (Wets and Bianchi [27], Wets et al. [29]); EpiRisk Research used them to build interest rates and currency exchange models, and this has now also been packaged by independent software providers. Section 2.2 provides an illustration. In the energy domain, epi-splines are used to compute day-ahead forecasts of electricity (load) demand (Feng et al. [13], Rios et al. [16]). In fact, there epi-splines are the main tools for building a stochastic process that provide not only a point forecast but conditional distributions of demand for a 24-hour period. The situation is similar in forecasting of commodity prices, where epi-splines result in remarkably accurate estimates

of copper prices (Wets and Rios [28]). Epi-splines are well suited for probability density estimation, especially in the context of little data (Casey and Wets [7], Royset and Wets [21], Sood and Wets [25]). Uncertainty quantification and simulation input and output analysis are applications that directly benefit from the epi-spline technology for density estimation (Royset et al. [22], Singham et al. [24]). Variograms are central tools in spatial statistics, and preliminary tests show the promise of epi-splines also in that area (Royset and Wets [20]).

Initial efforts on epi-splines centered on functions defined on a compact subset of the real line; see Royset and Wets [20] for an overview. Although this covers many important applications, higher dimensions undoubtedly arise too. Initial efforts in image reconstruction (Royset and Wets [19], Sood and Wets [25]), density estimation (Royset and Wets [19], Sood and Wets [25]), response surface building (Royset and Wets [19]), and graph models show the possibilities (§2.4). In this section, we provide a few examples that help motivate the analysis of the FIP, with additional ones furnished by §6.

2.1. Curve Fitting

An elementary example from curve fitting highlights central components of the FIP. Figure 1 illustrates 20 data points and plots a least-squares approximating fifth-degree polynomial. The polynomial fit is clearly too coarse, with polynomials of other degrees resulting in even worse approximations. This was recognized early and was a major motivation for the development of fitting procedures based on piecewise polynomials, specifically splines (Schoenberg [23]). We fit the data relying on four such procedures. Each one assumes, implicitly, that the actual function belongs to a particular family of functions and relies on a specific fitting criterion. The first three assume that the function is continuous, whereas the last one allows for a potential discontinuity at $x = 0.525$. We begin by relying on common cubic spline fitting. The solid line in Figure 2 plots the cubic interpolating spline with the usual Lagrangian end conditions, whereas the dashed line is the least-squares cubic smoothing spline with smoothing penalty $1 \cdot 10^{-4}$ assigned to the second-order derivative term. The fits improve over the polynomial fit, though the “smoothness” of the smoothing spline curve depends on the choice of penalty.

The final two procedures rely on epi-splines built from second-degree polynomials. Suppose that in addition to the data, the analyst believes that the actual function is continuously differentiable, with second-order derivatives in the range $[-100, 100]$. This external information forms constraints on the allowable functions that together with a least-squares criterion give the dashed line in Figure 3. The epi-spline line resembles the smoothing spline of Figure 2, with the former having a slightly better fit at some points. The smoothing spline is a result of minimizing squared error *plus* a weighted integral of second-order derivatives. The epi-spline

FIGURE 1. Curve fitting with a polynomial of degree 5.

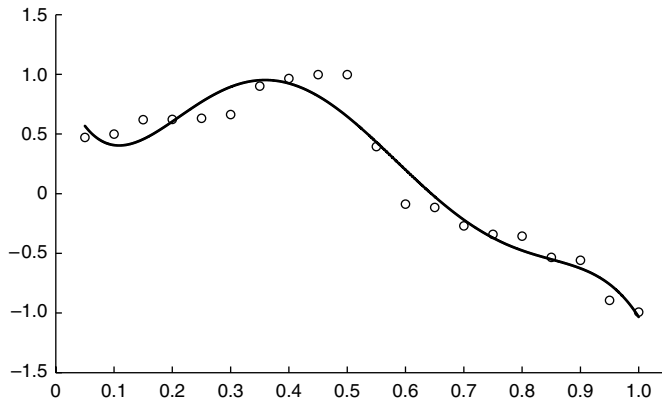
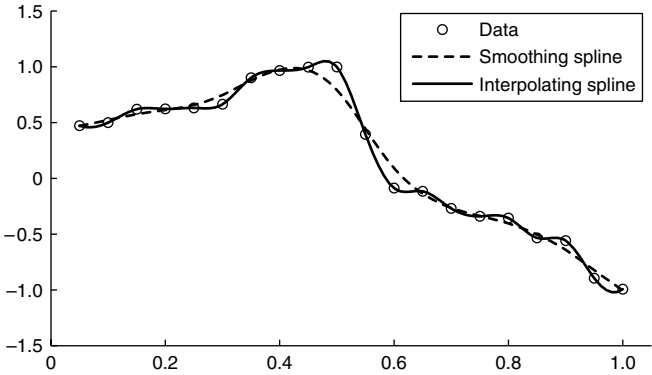


FIGURE 2. Cubic interpolating and smoothing spline fits.



simply minimizes the squared error but accounts for (pointwise) bounds on the second-order derivatives.

Suppose that external information further indicates that there may be a discontinuity around $x = 0.525$. Then an epi-spline illustrated by the solid line in Figure 3 yields what appears to be the best fit seen so far. Although not pursued further here, other epi-splines could be calculated that satisfy additional constraints derived from external information about concavity and other jumps, or by collecting further data points. A sequence of such problems can be thought of as approximations of an actual FIP involving an infinite number of data points and general functions. Below we discuss justifications for such approximations, especially those involving replacing general functions under consideration by epi-splines.

2.2. Financial Curves

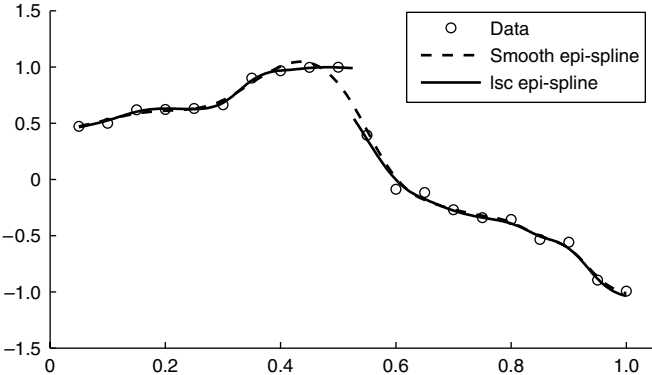
It is standard procedure in financial engineering to use a family of similar instruments consisting of Treasury bonds, annuities, commercial papers, and/or related assets to estimate their zero curves, including the spot rate curve, the forward rate curve, and the discount factor curve. These curves are intimately related as any one of them can be derived from knowing any one of the others; see Wets et al. [29], Wets and Bianchi [27], and references therein. Here, we consider the problem of finding a discount factor curve d for the sake of illustration.

Suppose that for each instrument $i = 1, 2, \dots, I$, there is a cash flow yielding

$$\text{cash payments } p_0^i, p_1^i, \dots, p_{N_i}^i \text{ at points in time } t_0^i, t_1^i, \dots, t_{N_i}^i, \text{ respectively.}$$

The first one of these “payments” usually corresponds to purchasing price, the remaining ones to the income generated; so, p_0^i is usually negative, whereas the remaining ones are positive.

FIGURE 3. Second-order epi-spline fits with and without discontinuity.



To simplify the notation, we let $t_0^i = 0$ (today) and $T \geq t_{N_i}^i$. For example, T could be related to the longest maturity time of all the instruments. Since one can reasonably assume that the discounted value of future payments balances the initial expense, or, more generally, the signed, discounted payment stream equals zero, identifying a discount factor curve amounts to finding a real-valued function d on $[0, T]$ with

$$\sum_{j=0}^{N_i} d(t_j^i) p_j^i = 0 \quad \text{for all } i = 1, 2, \dots, I.$$

External information about the nature of discount factor curves dictates further that $d(0) = 1$ (the value of money today is its face value), $d \geq 0$ (discount factors are nonnegative numbers), and the derivative $d' \leq 0$ (discount factors cannot increase with time). Additional information that relates to the desirable properties of the associated forward rates curves (see Wets and Bianchi [27, §1]) justifies the restrictions to smooth curves, and we therefore also require that d be continuously differentiable. Since ensuring that all of the above I equations are satisfied may not be possible because of inconsistent pricing in the markets, we formulate the problem in terms of a norm-minimizing optimization problem:

$$\text{minimize } \left\| \left(\sum_{j=0}^{N_1} d(t_j^1) p_j^1, \dots, \sum_{j=0}^{N_I} d(t_j^I) p_j^I \right) \right\| \quad \text{such that } d(0) = 1, d \geq 0, d' \leq 0,$$

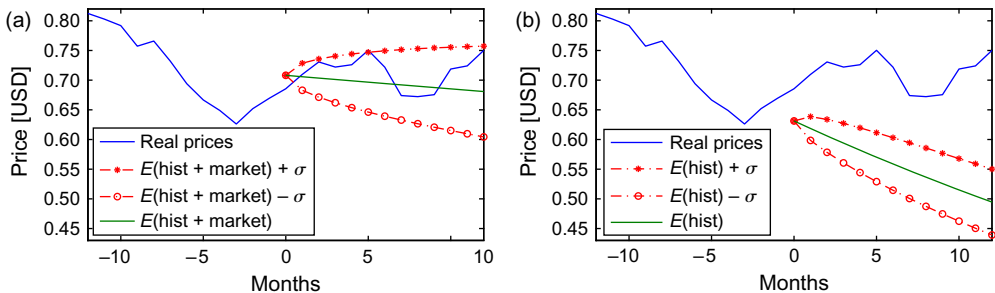
where it is implicitly assumed that d is a continuously differentiable function, and $\|\cdot\|$ is a well-chosen norm. The problem is clearly an FIP, with constraints derived from insight about the situation.

2.3. Commodity Price Forecasts

A more complicated example arises in commodity pricing. In Wets and Rios [28, §3], epi-spline technology builds a stochastic process for short-term copper prices, say, for the next 6–12 months. The effort relies on a common model for the evolution of copper prices as well as closely related quantities such as interest rates and exchange rates based on geometric Brownian motion. The drift and volatility terms in the corresponding stochastic differential equation, however, are estimated using epi-splines. The available information is historical prices; see, for example, the line in Figure 4(a) that stretches across the figure, where “today” is month 0 so that prices only up to that time are available. Moreover, we also know current prices of future contracts for copper deliveries as well as for various interest rate and currency exchange instruments.

Epi-spline technology enters then twice. First, prices of future contracts are converted into (implied future) spot prices. This is achieved by computing a discount factor curve in the manner described above and then converting it into a spot rate curve using standard

FIGURE 4. Epi-spline estimates of copper prices with (a) and without (b) future contract information.



expressions. Second, the historical prices, today's price, and the spot market prices just computed for, say, the next nine months provide data that are fitted using a second-order epi-spline under the additional external information that it is continuously differentiable. The resulting fit provides estimated drift terms for the stochastic differential equation. The solid line in Figure 4(a) starting at month 0 shows the price drift for the next 12 months. We note that in this approach the initial condition of the process is not today's observed price. A justification for proceeding in this fashion is that one should view today's (observed) spot price as the "actual" spot price perturbed by some noise; empirical calculations carried out in Wets and Rios [28, §6.1] confirm that this approach yields much better results.

Finally, the volatility parameters of the process are obtained by a least-squares fit of the covariance after adjusting the observed historical prices and the forecasted spot prices by subtracting their estimated expectation; for more about this approach, see Wets and Rios [28]. Figure 4(a) illustrates the volatility with dotted lines corresponding to the 95% confidence interval for the process at each point in time. Figure 4(b) provides similar results, but without using epi-spline technology to convert information about future contracts into implied future spot prices. It is apparent that including this additional external information dramatically improves the quality of the estimates.

2.4. Density Estimation on Graphs

A final illustration is provided by a problem in graph theory. There is a rapidly growing literature on models for graphs that rely on exponential families of probability density functions

$$h_{\beta}(G) = \exp\left(\sum_{i=1}^I \beta_i T_i(G) + c_{\beta}\right)$$

to describe the likelihood of a graph G characterized by its number of edges ($T_1(G)$), triangles ($T_2(G)$), stars ($T_3(G)$), cycles ($T_4(G)$), etc., with $\beta = (\beta_1, \beta_2, \dots, \beta_I)$ being a vector of parameters and c_{β} a normalization coefficient; see Chatterjee and Diaconis [8] and references therein. From observed data in the form of a collection of graphs, there are significant challenges to estimate the parameters β . In Chatterjee and Diaconis [8], we find an approach that relies on graph limits in the sense of Lovász and large-deviation results for random graphs, and leads to an infinite-dimensional problem whose optimal value is an estimate of the crucial normalization constant c_{β} .

Specifically, following Chatterjee and Diaconis [8], we define

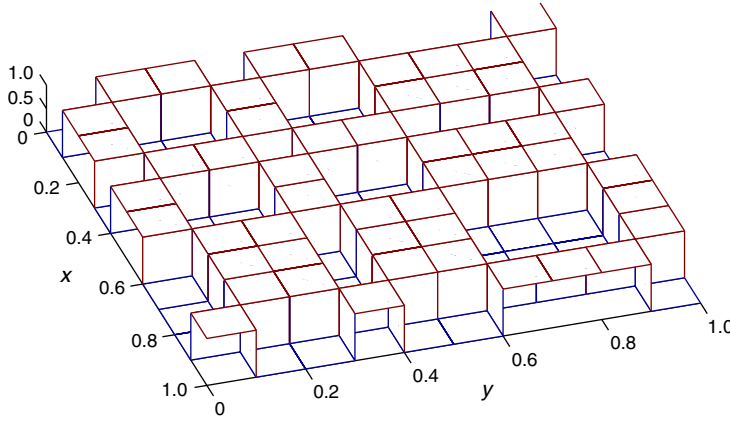
$$\begin{aligned}\psi_0(f) &= \int_0^1 \int_0^1 \left[\frac{1}{2} f(x_1, x_2) \log f(x_1, x_2) + \frac{1}{2} (1 - f(x_1, x_2)) \log(1 - f(x_1, x_2)) \right] dx_1 dx_2; \\ \psi_1(f) &= \int_0^1 \int_0^1 f(x_1, x_2) dx_1 dx_2; \\ \psi_2(f) &= \int_0^1 \int_0^1 \int_0^1 f(x_1, x_2) f(x_2, x_3) f(x_1, x_3) dx_1 dx_2 dx_3.\end{aligned}$$

The infinite-dimensional problem is then to

$$\text{minimize } \psi_0(f) - \beta_1 \psi_1(f) - \beta_2 \psi_2(f) \quad \text{such that } f \text{ is symmetric and } 0 \leq f \leq 1$$

over all measurable functions on $[0, 1]^2$. Here, symmetric means that $f(x_1, x_2) = f(x_2, x_1)$ for all $(x_1, x_2) \in [0, 1]^2$. An approach relying on epi-splines for solving this problem is currently under investigation. Figure 5 provides an example of a minimizing function f for the choice $\beta_1 = 4$ and $\beta_2 = -7$. This problem only indirectly relates to "fitting data" and shows the breadth of possible applications of the epi-spline technology.

FIGURE 5. Example of minimizing function f in a graph density problem.



3. Problem Formulations

As indicated above, the FIP is an optimization problem, and we now turn to its formulation. In particular, we make distinction between an actual, conceptual problem and an approximate problem to be solved computationally. The section ends with formulation of constraints derived from external information and practical guidelines.

3.1. Actual Problem

We formulate the FIP as the problem of determining a function f defined on $\mathbb{R}^n$ that minimizes some criterion given by a functional ψ while also satisfying constraints given by a feasible set F . The problem may be conceptual, involving information not available and requiring operations not implementable during its solution. This “actual problem” therefore takes the form

$$\text{Actual problem (P): } \min \psi(f) \quad \text{such that } f \in F \subseteq \mathcal{F},$$

where $\mathcal{F}$ is a space of functions. The criterion functional ψ could be, for example,

$$\begin{aligned} \psi(f) &= \sum_{j=1}^J (y^j - f(x^j))^2, \\ \psi(f) &= \sum_{j=1}^J |y^j - f(x^j)|, \\ \psi(f) &= \max_{j=1, \dots, J} |y^j - f(x^j)|, \end{aligned}$$

where $\{(x^j, y^j)\}_{j=1}^J$ are observed data. These terms express errors from the observed data as measured by mean squared, mean absolute, and max absolute errors, respectively. More generally, for a known reference function f^0 , we may have

$$\psi(f) = \|f - f^0\|,$$

where $\|\cdot\|$ is a norm defined on the space of functions under consideration. For random variables X and Y , with X n -dimensional, expressing possible “future” data, examples of criteria are

$$\begin{aligned} \psi(f) &= E[f(X)], \\ \psi(f) &= E[\max\{0, Y - f(X)\}]/(1 - \alpha) + E[Y - f(X)], \\ \psi(f) &= \mathcal{E}(Y - f(X)). \end{aligned}$$

The first example focuses on the expected value of the random variable $f(X)$ as arises in density estimation using log-likelihood maximization and an exponential transformation; see Royset and Wets [21]. The second expression states the Koenker–Bassett error used in quantile regression, with $\alpha \in [0, 1)$. The third gives the error between random variables Y and $f(X)$ as expressed by an error measure, of which the Koenker–Bassett and mean squared errors are special cases; see Rockafellar and Uryasev [17] for other examples that lead to many different types of generalized regression. Two additional examples are furnished by

$$\begin{aligned}\psi(f) &= \int \|f''(x)\|^2 dx, \\ \psi(f) &= \min_x f(x),\end{aligned}$$

which do not directly relate to the data. The first one is common in applications such as interpolation. Here, the criterion is to obtain a function with minimum “variation” as measured by the integrated squared value of the second-order derivative, or, more generally, a squared norm of the Hessian matrix. The last example illustrates the versatility of the framework as it allows for criteria functionals that themselves involve optimization.

The feasible region F likewise comes in a large variety of forms. The restrictions to functions that are smooth, convex, monotone, and log-concave¹ naturally arise in applications. For example, an analyst may be rather certain that increasing input results in increasing output in a regression analysis. Consequently, she should restrict the consideration to increasing functions. The prevalence of “normal-looking” probability density functions in an area may convince an analyst to consider only log-concave densities, which have only one mode. However, the possibilities extend much further. The functions of interest in an application may be known to attain specific values at a finite number of points, or be zero, $-\infty$, or ∞ outside a certain (compact) set. The functions could be nonsmooth, but with subgradients that are contained in certain sets and possess specific properties, or could be within some distance from a reference function. In fact, any one application may need to contend with many or all of this external information at the same time. At the end of this section, we give some concrete examples of how these pieces of information can be formulated mathematically within an epi-spline framework. We refer, for example, to Royset and Wets [21] for illustrations of the effect of external information in improving estimates.

The distinction between a criterion functional and constraints defining the feasible set is less important than it first appears. It is well known in practical application of optimization models and methods that an objective function can be converted into a constraint and a constraint function can take the place of an objective function with computational and modeling benefits. However, some requirements such as continuity, convexity, and differentiability are most naturally formulated as constraints. Quantifications of “goodness of fit” lend themselves as criteria functionals. Typically, application-specific conditions indicate the most beneficial formulation, with the opportunity to include arbitrary constraints in (P) offering significant flexibility.

Two classical examples provide additional context. A problem of smooth interpolation aims to determine a function f defined on $\mathbb{R}$ that

$$\text{minimizes } \int f''(x)^2 dx \quad \text{such that } f(x^j) = y^j \text{ for all } j = 1, 2, \dots, J,$$

i.e., to determine the “smoothest” function that interpolates the points (x^j, y^j) , $j = 1, 2, \dots, J$. When restricted to functions in a certain Sobolev space, the unique solution of this problem is a cubic spline² (Holladay [15]; see also Atteia [1], de Boer [11] for numerous extensions).

¹ We recall that a function f is log-concave if $\log f$ is concave.

² Cubic splines are piecewise polynomial functions of third degree that are twice continuously differentiable.

The solid line in Figure 2 illustrates such a cubic spline for the data set of §2.1. Cubic smoothing splines balance interpolation of the data with “smoothness” of the curve and arise as optimal solutions of

$$\min \left\{ \sum_{j=1}^J (y^j - f(x^j))^2 + \lambda \int f''(x)^2 dx \right\},$$

where $\lambda \geq 0$ is a smoothing penalty (Egerstedt and Martin [12], Wahba [26]). The dashed line in Figure 2 provides an illustration for $\lambda = 1 \cdot 10^{-4}$. Other “regularizations” by ℓ_1 - and ℓ_0 -norms lead to similar problems. It is clear that these examples are special cases of the actual problem (P).

In the interpolation and smoothing spline examples, optimal solutions are automatically cubic splines, and the choice of function space to consider falls naturally to a Sobolev space to ensure the existence and integrability of second-order derivatives. The general case (P) requires more care to ensure a sufficiently rich class of functions that can capture (essentially) all possibilities arising in practice. We would like to allow for discontinuous functions; functions with arbitrary rates of growth—for example, unbounded derivatives; and functions that assume the values $-\infty$ and ∞ . Discontinuities certainly arise in the nonparametric estimation of densities, in response surface building, and in curve fitting. Curve fitting may also lead to functions that grow too fast to be integrable in some sense. The possibility of function values $-\infty$ and ∞ may be needed when implicit constraints are present, for example, in curve fitting of functions with an effective domain less than the whole domain; it may also arise in applications involving rescaling, for instance, exponential rescaling, where zero values need to be allowed. Consequently, we let the space of function under consideration, $\mathcal{F}$, be the set, or possibly a subset, of lower semicontinuous (lsc) functions from $\mathbb{R}^n$ to the extended real line $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, excluding the trivial function $f \equiv \infty$, which is identical to infinity everywhere. The left portion of Figure 6 illustrates an lsc function on $\mathbb{R}$. As shown, an lsc function may be *extended real valued* with values $-\infty$ and ∞ and have discontinuities, at which point the function “jumps down.” The formal definition of an lsc function requires additional notation. For any f on $\mathbb{R}^n$, we let

$$\liminf_{x' \rightarrow x} f(x') := \lim_{\delta \downarrow 0} \left[\inf_{x' \in \mathbb{B}(x, \delta)} f(x') \right],$$

where $\mathbb{B}(x, \delta) := \{x' \in \mathbb{R}^n \mid \|x' - x\| \leq \delta\}$ is the ball of radius δ centered at x . Informally, we think of $\liminf_{x' \rightarrow x} f(x')$ as the smallest value of f “near” x . Then,

$$\text{a function } f \text{ is lsc at a point } x \quad \text{if } \liminf_{x' \rightarrow x} f(x') \geq f(x).$$

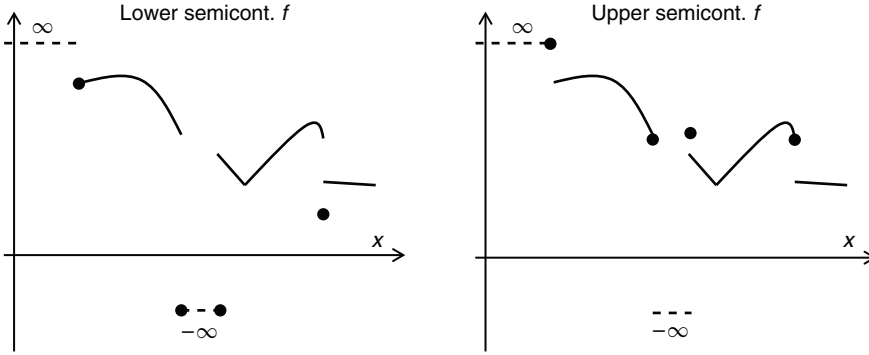
A function is lsc if it is lsc for all $x \in \mathbb{R}^n$. We denote by $\text{lsc-fcns}(\mathbb{R}^n)$ this set of lsc functions, excluding $f \equiv \infty$. The lsc functions are complemented by upper semicontinuous functions that “jump up” at points of discontinuity, as illustrated by the right portion of Figure 6. The definition is similar. In fact, the lsc functions yield the upper semicontinuous functions after a sign change. A parallel development with such functions is therefore mostly superfluous.

Since external information may indicate that the consideration of, for example, continuous or continuously differentiable functions suffices, not all applications require the full generality of $\text{lsc-fcns}(\mathbb{R}^n)$. We therefore let $\mathcal{F}$ in the actual problem be equal to $\text{lsc-fcns}(\mathbb{R}^n)$ or an appropriate subset. The choice of representing external information by the constraint set F , or, alternatively, in the definition of the space of functions $\mathcal{F}$, is purely a technicality.

3.2. Approximations

The actual problem (P) rarely admits closed-form solutions, and computational strategies become essential. However, optimization over functions instead of over a finite number of

FIGURE 6. Lower (left) and upper (right) semicontinuous functions.



parameters is not implementable in a computational method. The criterion functional may also involve integrals, and the number of constraints could be infinite, both requiring some sort of approximation. The actual problem could also represent an “ideal” situation with full information that should be contrasted with a problem involving only observed data. Regardless of the situation, there is a need to consider approximations of (P) . We start by defining approximations of lsc functions in terms of simpler functions called epi-splines that involve only a finite number of parameters.

Epi-splines are defined on $\mathbb{R}^n$, but we first present the one-dimensional case for simplicity.

One-dimensional epi-splines on a mesh. A *mesh* is a finite number of points

$$-\infty = m_0 < m_1 < m_2 < \cdots < m_{N-1} < m_N = \infty$$

on the real line as illustrated by Figure 7. An *epi-spline*³ of order p on such a mesh is a real-valued function s such that

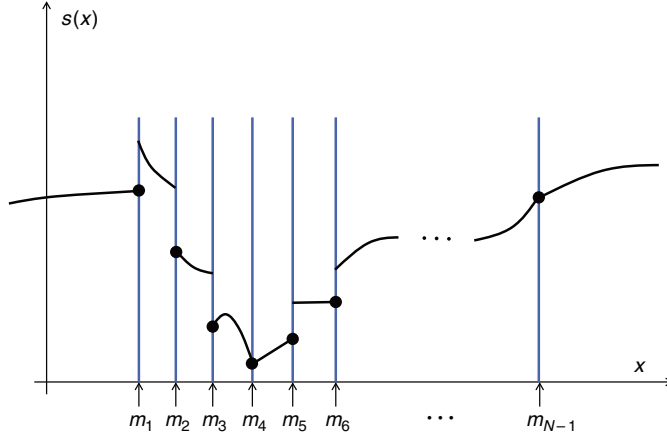
on each open interval (m_{k-1}, m_k) , $k = 1, \dots, N$, s is polynomial of degree p , and
for every $x \in \mathbb{R}$, $s(x) = \min\{\lim_{t \downarrow 0} s(x+t), \lim_{t \downarrow 0} s(x-t)\}$.

Figure 7 shows an epi-spline of the second order, i.e., $p = 2$, which on each open interval (m_{k-1}, m_k) is a second-degree polynomial. An epi-spline of order $p = 0$ is a piecewise constant function, and one of order $p = 1$ is a piecewise affine function with possibilities of jumps at the mesh points $m_1, m_2, \dots, m_{N-1}$. The second condition ensures that the value $s(x)$ is the smaller of the left and right limits of s at x , which, of course, exist since s is continuous (polynomial) except at the mesh points. This ensures that s is lsc.

Epi-splines are structurally related to splines but are more flexible. In fact, interest in splines of order p with less than the standard $(p-1)$ -times continuous differentiability at their “knots” goes back to Curry and Schoenberg [9, 10]. However, mesh points $m_1, \dots, m_{N-1}$ should not be identified with the knots of such splines since no assumptions are made about information being available at mesh points, and no continuity requirements are placed on epi-splines at (precisely) these mesh points. In fact, the partition can be selected completely freely. Continuity and any other condition can still be ensured when needed through constraints in (P) , as discussed at the end of this section. Since from an etymological viewpoint the (Greek) prefix “epi” can be interpreted as meaning “higher” or “better,” the term “epi-splines” is certainly appropriate for these more “general” splines. In §§4 and 5, we see that the analysis of epi-splines relies heavily on epi-convergence and the epi-topology, which provides further justification for the name.

³ In Royset and Wets [19], these epi-splines are called lsc epi-splines, but here we exclusively focus on such epi-splines, and the prefix is therefore superfluous. Basic epi-splines defined only on a compact interval of $\mathbb{R}$ are given in Royset and Wets [20].

FIGURE 7. Mesh and epi-spline on $\mathbb{R}$.



Higher-dimensional epi-splines. Epi-splines in higher dimensions are defined similarly but require additional notation. We denote by $\text{cl } S$ the closure of a subset S of a Euclidean space. A finite collection $R_1, R_2, \dots, R_N$ of open subsets of $\mathbb{R}^n$ is a *partition* of $\mathbb{R}^n$ if

$$\bigcup_{k=1}^N \text{cl } R_k = \mathbb{R}^n \quad \text{and} \quad R_k \cap R_l = \emptyset \quad \text{for all } k \neq l.$$

There are clearly many possible partitions, but, in practice, those consisting of rectangular boxes and simplexes appear to suffice. We adopt a “total degree” convention and say that a polynomial in n dimensions is of *total degree* p if it is expressed as a finite sum of polynomial terms, each having the sum of the powers of the variables being no larger than p . We note that the total number of terms in such a polynomial is at most $(n+p)!/(n!p!)$. Another convention would have had only minor consequences for the following exposition.

An epi-spline s of order p defined on $\mathbb{R}^n$ with partition $\mathcal{R} = \{R_k\}_{k=1}^N$ is a real-valued function that

$$\begin{aligned} &\text{on each } R_k, k = 1, \dots, N, s \text{ is polynomial of total degree } p, \text{ and} \\ &\text{for every } x \in \mathbb{R}^n, s(x) = \liminf_{x' \rightarrow x} s(x'). \end{aligned}$$

As in the one-dimensional case, an epi-spline is polynomial on open sets. The second condition ensures that it is also lsc by defining the value at boundary points of the partition appropriately. Figure 8 illustrates an epi-spline of order 2 on $\mathbb{R}^2$. The partition consists of rectangles. The family of all epi-splines of order p on $\mathbb{R}^n$ with partition $\mathcal{R}$ is denoted by $\text{e-spl}_n^p(\mathcal{R})$. Since an $s \in \text{e-spl}_n^p(\mathcal{R})$, with $\mathcal{R} = \{R_k\}_{k=1}^N$, involves N polynomials of total order p , it is fully characterized by

$$n_e := N(n+p)!/(n!p!) \quad (1)$$

parameters, which may be large but finite.

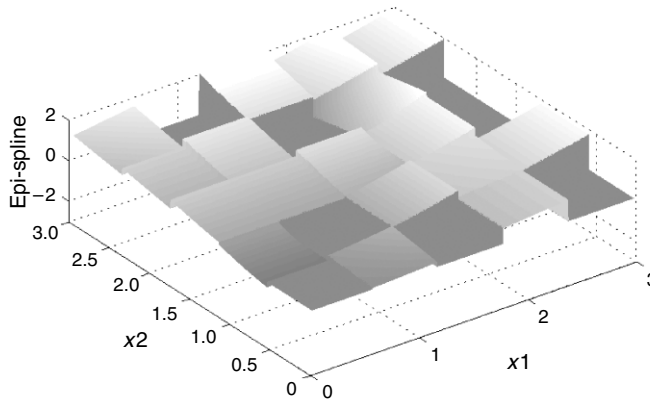
Approximate problem. To also allow for the possibility of approximations in the criterion functional ψ and the feasible set F , we define an *approximate criterion* ψ^ν and an *approximate feasible set* $F^\nu \subset \mathcal{F}$. Relying on epi-splines, we define the approximate functions under consideration to be

$$\mathcal{S}^\nu \subseteq \text{e-spl}_n^p(\mathcal{R}^\nu) \cap \mathcal{F}.$$

The superscript $\nu = 1, 2, \dots$ indicates that a family of approximations can be considered, possibly with gradually more refined partition $\mathcal{R}^\nu$, for example. These approximations lead to an

$$\text{Approximate problem } (P^\nu): \quad \min \psi^\nu(s) \quad \text{such that } s \in F^\nu \cap \mathcal{S}^\nu$$

that approximates the actual problem (P) .

FIGURE 8. Epi-spline on $\mathbb{R}^2$.

(P^ν) is an optimization problem over functions, but they are all epi-splines and therefore representable by a finite number of parameters. Consequently, (P^ν) is a *finite-dimensional* optimization problem that is solved by optimizing over the $N^\nu(n + p^\nu)!/(n!p^\nu!)$ (real) parameters describing the family of epi-splines under consideration. Here, N^ν is the number of open sets in the partition $\mathcal{R}^\nu$ of $\mathbb{R}^n$. We refer to an optimization problem over such parameters as a *parametric optimization problem* corresponding to (P^ν) ; see §5 for further discussion. Under the assumption that additional complications are removed by approximations of the criterion functional and feasible set, the solution of parametric optimization problems can be obtained by standard algorithms, or possibly more efficiently by specialized algorithms in certain cases.

With the introduction of approximations, we need to address the possibility that solutions of approximate problems (P^ν) may not relate to those of the actual problem (P) . Under what conditions will solutions of (P^ν) be approximate solutions of (P) ? More fundamentally, can we approximate any lsc function by epi-splines? These questions require the discussion of approximation theory of optimization problems as reviewed in §4, with answers provided in §5. We end this section, however, with examples of external information and practical guidelines.

3.3. External Information

External information needs to be converted into mathematical forms and included as constraints in the parametric optimization problem corresponding to (P^ν) . The formulation of such constraints depends on the representation of the polynomials in an epi-spline. Of course, there are many possible representations, some of which are known to be numerically more beneficial than others. Here, we only consider one-dimensional epi-splines and the simple representation

$$q^k(x) = a_0^k + a_1^k x + a_2^k x^2 + \cdots + a_p^k x^p, \quad x \in \mathbb{R}, \quad (2)$$

of the polynomial describing an epi-spline on the interval (m_{k-1}, m_k) , $k = 1, 2, \dots, N$. With N intervals partitioning $\mathbb{R}$, the epi-spline is defined by N such polynomials with a total of $N(p+1)$ parameters. We next give some examples of external information and the formulation of corresponding constraints using this representation; see Royset and Wets [19] for extensions to higher dimensions and Royset and Wets [21] for constraints related to density estimation.

Continuity. An epi-spline is continuous if the polynomials $q^1, q^2, \dots, q^N$ defining the epi-spline coincide at the mesh points $m_1, m_2, \dots, m_{N-1}$. This is easily achieved by the constraints

$$a_0^k + a_1^k m_k + a_2^k m_k^2 + \cdots + a_p^k m_k^p = a_0^{k+1} + a_1^{k+1} m_k + a_2^{k+1} m_k^2 + \cdots + a_p^{k+1} m_k^p, \quad k = 1, 2, \dots, N-1.$$

Of course, this constraint could be enforced only at a subset of the mesh points to allow for discontinuities in some areas, but not in others as done in Figure 2.

Continuous differentiability. An epi-spline is continuously differentiable if it is continuous, it is ensured by the above equations, and the derivatives of the polynomials $q^1, q^2, \dots, q^N$ coincide at the mesh points $m_1, m_2, \dots, m_{N-1}$. This is achieved by the constraints

$$a_1^k + 2a_2^k m_k + \dots + pa_p^k m_k^{p-1} = a_1^{k+1} + 2a_2^{k+1} m_k + \dots + pa_p^{k+1} m_k^{p-1}, \quad k = 1, 2, \dots, N-1,$$

where we assume that $p \geq 1$. Continuous differentiability is automatic for a continuous epi-spline of order $p = 0$.

Fixed values. We ensure that an epi-spline satisfies the function value $s(x)$ on a set $S_k \subset (m_{k-1}, m_k)$ by the constraints

$$a_0^k + a_1^k x + a_2^k x^2 + \dots + a_p^k x^p = s(x), \quad x \in S_k.$$

This may at first give the impression that an infinite number of constraints is needed. However, if S_k has more than $p+1$ distinct points, then it suffices to select $p+1$ distinct points at which to enforce the constraints. This follows from the fact that a polynomial of degree p is uniquely defined by its value at $p+1$ points. The equality can naturally be replaced by inequality if only bounds are available. Moreover, if the values of an epi-spline are of no interest beyond a certain mesh point, then the polynomials defining the epi-spline beyond that point are of course immaterial and can be ignored.

Monotonicity. We ensure that an epi-spline of order $p \geq 1$ is nonincreasing by the constraints

$$\begin{aligned} a_1^k + 2a_2^k x + \dots + pa_p^k x^{p-1} &\leq 0, \quad x \in (m_{k-1}, m_k), k = 1, 2, \dots, N, \quad \text{and} \\ a_0^k + a_1^k m_k + a_2^k m_k^2 + \dots + a_p^k m_k^p &\geq a_0^{k+1} + a_1^{k+1} m_k + a_2^{k+1} m_k^2 + \dots + a_p^{k+1} m_k^p, \quad k = 1, 2, \dots, N-1. \end{aligned}$$

The first set of constraints ensures that the polynomials making up an epi-spline are nonincreasing on the intervals on which they define the epi-spline. The second set imposes the restriction that when moving from left to right, the epi-spline must not jump up at mesh points. The second requirement is automatically satisfied for continuous epi-splines. In the case of $p = 2$, the first condition simplifies to the two constraints

$$a_1^k + 2a_2^k m_{k-1} \leq 0 \quad \text{and} \quad a_1^k + 2a_2^k m_k \leq 0,$$

as, in that case, the derivative of the epi-spline is simply linear. For any p , the first and the second conditions hold if

$$a_1^k + 2a_2^k m_k + \dots + pa_p^k m_k^{p-1} \leq 0$$

under the additional assumption that the epi-spline is convex, since then the epi-spline is continuous with a nondecreasing (left/right-sided) derivative.

Convexity. We ensure that an epi-spline is convex by the continuity constraints above and

$$\begin{aligned} 2a_2^k + \dots + p(p-1)a_p^k x^{p-2} &\geq 0, \quad x \in (m_{k-1}, m_k), k = 1, 2, \dots, N, \quad \text{and} \\ a_1^k m_k + 2a_2^k m_k + \dots + pa_p^k m_k^{p-1} &\leq a_1^{k+1} m_k + 2a_2^{k+1} m_k + \dots + pa_p^{k+1} m_k^{p-1}, \quad k = 1, 2, \dots, N-1. \end{aligned}$$

The first set of constraints ensures that the second-order derivative of the polynomials constituting an epi-spline are nonnegative and therefore convex on the intervals on which they define the epi-spline. For $p < 2$, the constraints are superfluous. The second set imposes the restriction that when moving from left to right, the derivative of the epi-spline must not jump down at mesh points. The second requirement is automatically satisfied for continuously differentiable epi-splines and can be ignored for the case $p = 0$. If an epi-spline is of the second

order, then the first set of constraints simplifies to $a_2^k \geq 0$ for $k = 1, \dots, N$. If $p = 3$, the first set of constraints simplifies to

$$2a_2^k + 6a_3^k m_{k-1} \geq 0 \quad \text{and} \quad 2a_2^k + 6a_3^k m_k \geq 0$$

because then the second-order derivative is simply linear.

Log-concavity and modality. We recall that the composite function $h(x) = e^{-s(x)}$, $x \in \mathbb{R}^n$ is log-concave if and only if s is convex. Consequently, it can be beneficial to use such an exponential transformation of an epi-spline when log-concavity is desirable as in the case of probability density estimation; see Royset and Wets [21]. In this case, the convexity constraints above would ensure log-concavity. Log-concavity implies unimodality. However, the converse is not true. It is more complicated to ensure unimodality and not necessarily log-concavity. In the case of a continuous function, one could ensure unimodality by designating one mesh point $m_{k'}$ as the mode, and then constraining the function to be increasing and decreasing on $(m_{k'-1}, m_{k'})$ and $(m_{k'}, m_{k'+1})$, respectively, and nondecreasing on $[m_0, m_{k'-1})$ and nonincreasing on $(m_{k'+1}, m_N]$. Solving the resulting problem gives a unimodal candidate function. The process is repeated for alternative mode locations m_k , $k = 0, 1, \dots, N$, $k \neq k'$, and the function with the best criterion functional is retained as the optimal function. K -modality is achieved similarly by partitioning $[m_0, m_N]$ into K intervals, with each having a unimodal constraint. The process must be repeated for each partition of interest. To specify that certain m_k are modes is achieved by ensuring that the function is increasing and decreasing on (m_{k-1}, m_k) and (m_k, m_{k+1}) , respectively.

We note that the above constraints, and many similar ones, are mostly *linear* in the parameters describing an epi-spline. Consequently, their inclusion in a parametric optimization problem corresponding to (P^ν) requires little additional computational effort.

3.4. Implementation Guidelines

The formulation and implementation of an approximate problem (P^ν) in a specific context require the choice of order p and partition $\mathcal{R}$. The order is usually easy to choose, as the extensive experience with classical polynomial splines over more than half a century strongly indicates that low-order splines are preferred to higher-order ones because they avoid the oscillatory behavior of high-degree polynomials. Typically, order 2 or 3 is recommended. However, the consideration of derivatives might dictate slightly higher orders. For example, external information about a k th derivative of the actual function dictates that at least order k or maybe $k + 1$ needs to be used; see Royset and Wets [19, 20] for further discussion.

The choice of mesh (in one dimension) is usually straightforward; one should, in view of the following theory, select a mesh as fine as is computationally possible. However, when computing times are important or external information indicates that the actual function is “nearly polynomial,” one could consider coarse meshes. The exact location of mesh points could be dictated by external information, for example, about points of discontinuity. When no relevant information is available, a choice with evenly dispersed mesh points is natural. The situation in higher dimensions is more complex with the multitude of possible partitions. Unless special external information is available, convenience suggests polyhedral partitions, especially boxes and simplexes. In particular, simplex partitions appear versatile because they are compatible with continuous epi-splines in all dimensions even for order $p = 1$; see Royset and Wets [19].

Another issue is the representation of polynomials, with (2) being only one possibility. In fact, we know that this representation leads to poorly conditioned problems in certain contexts of interpolation and fitting. Although we have found no need for going beyond this representation, or one that is normalized with respect to mesh point locations, in the applications examined thus far, we foresee that efficiencies will accrue from further studies of this area, especially in higher dimensions.

4. Background on Epi-Convergence

The examination of epi-splines and the relationship between the actual problem (P) and approximate problems (P^ν) rely on variational analysis and, especially, the approximation theory of optimization problems with the notion of epi-convergence taking center stage.

In this context, it is not possible to rely on the standard pointwise and uniform convergence notions. Pointwise does not guarantee the convergence of minimizers, e.g., $f^\nu \equiv 1$ on $[0, 1]$ except for $f^\nu(1/\nu) = 0$ converges pointwise to $f \equiv 1$ on $[0, 1]$, but clearly the minimizers of the f^ν do not converge to the set of minimizers of f , which is $[0, 1]$. On the other hand, uniform convergence will guarantee the convergence of the minimizers but can, at best, only be satisfied when the constraint set is not affected by the approximation(s), e.g., $f^\nu \equiv 0$ on $[0, 1 - 1/\nu]$ does not converge uniformly to $f \equiv 0$ on $[0, 1]$. One can view epi-convergence as a one-sided (unilateral) uniform convergence notion that can handle the challenges posed by an optimization problem and guarantees, in a way to be made precise a bit later, the convergence of the minimizers. Here, we provide the essential components; see Rockafellar and Wets [18] and Attouch and Wets [3] for a more comprehensive treatment.

We view a function on $\mathbb{R}^n$ through its *epi-graph*

$$\text{epi } f := \{(x, x_0) \in \mathbb{R}^{n+1} \mid f(x) \leq x_0\},$$

i.e., the set of points in $\mathbb{R}^{n+1}$ that lies no lower than the graph of f , as illustrated in Figure 9. We note that

$$f \in \text{lsc-fcns}(\mathbb{R}^n) \quad \text{if and only if } \text{epi } f \text{ is a nonempty closed subset of } \mathbb{R}^{n+1}.$$

Distances between such functions can therefore be based on distances between (closed) sets in $\mathbb{R}^{n+1}$.

We start with notation and let $\rho\mathbb{B} = \mathbb{B}(0, \rho)$ be the closed ball in a Euclidean space centered at the origin with radius $\rho \geq 0$ and

$$d(y, S) := \inf_{y' \in S} \|y - y'\|$$

be the standard distance between a point y and a subset S of a Euclidean space. We use the notation $A + B = \{a + b \mid a \in A, b \in B\}$ for the Minkowski sum of two sets A and B in a Euclidean space.

The classical Pompeiu–Hausdorff distance between sets $C, D \subset \mathbb{R}^n$,

$$d_\infty(C, D) := \sup_{y \in \mathbb{R}^n} |d(y, C) - d(y, D)| = \inf\{\eta \geq 0 \mid C \subset D + \eta\mathbb{B}, D \subset C + \eta\mathbb{B}\},$$

is not useful in the present context because epi-graphs are unbounded sets, and this distance is easily infinity. We therefore turn to two distances that depart from the Pompeiu–Hausdorff distance in different directions (see Attouch et al. [5] and Rockafellar and Wets [18, §4.H]).

FIGURE 9. Epi-graph of a function f on $\mathbb{R}$.

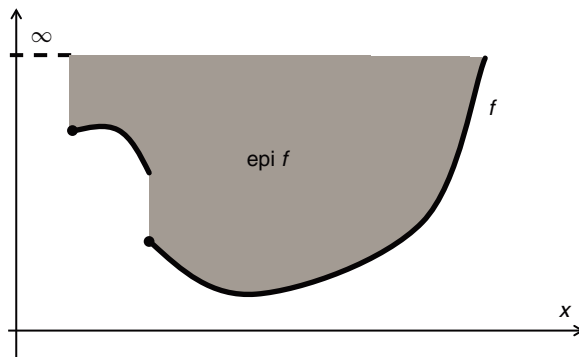
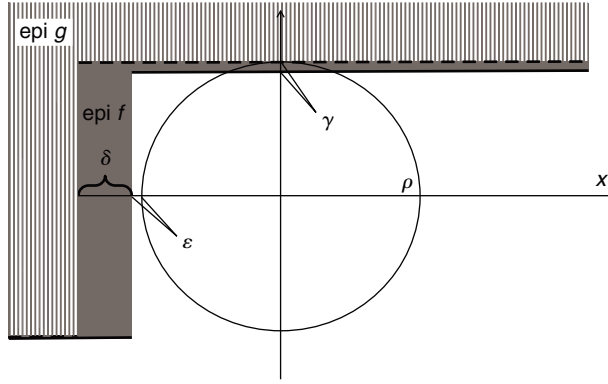


FIGURE 10. Illustration of distances: $d_\rho(f, g) = \delta$ and $\hat{d}_\rho(f, g) = \gamma$.

Focusing now on epi-graphs, we first define the ρ -epi-distance between two functions f and g on $\mathbb{R}^n$, which is given by a restriction of the first formula for the Pompeiu–Hausdorff distance:

$$d_\rho(f, g) = \sup_{\bar{x} \in \rho\mathbb{B}} |d(\bar{x}, \text{epi } f) - d(\bar{x}, \text{epi } g)|, \quad \rho \geq 0.$$

Obviously, $d_\rho(f, g) = d_\infty(\text{epi } f, \text{epi } g)$ for $\rho = \infty$. As illustrated in Figure 10, $d_\rho(f, g)$ gives the largest difference in distance between a point $\bar{x} = (x, x_0) \in \rho\mathbb{B} \subset \mathbb{R}^{n+1}$ and the respective epi-graphs. In this figure, with $f(x) = \rho - \gamma$ for $x > -\rho - \varepsilon$ and $f(x) = -\rho$ for $x \leq -\rho - \varepsilon$, $g(x) = \rho$ for $x > -\rho - \delta - \varepsilon$ and $g(x) = -\rho$ for $x \leq -\rho - \delta - \varepsilon$, and $\delta + \varepsilon < \rho$ and ε, γ small, the point $\bar{x} = (-\rho, 0)$ provides the maximizing point, with $d(\bar{x}, \text{epi } f) = \varepsilon$ and $d(\bar{x}, \text{epi } g) = \delta + \varepsilon$. Consequently, $d_\rho(f, g) = \delta$.

The second distance, the ρ -epi-distance estimate, relies on the second formula for the Pompeiu–Hausdorff distance but limits the focus to the ball centered at the origin with radius ρ :

$$\hat{d}_\rho(f, g) = \inf \{ \eta \geq 0 \mid \text{epi } f \cap \rho\mathbb{B} \subset \text{epi } g + \eta\mathbb{B}, \text{epi } g \cap \rho\mathbb{B} \subset \text{epi } f + \eta\mathbb{B} \}, \quad \rho \geq 0.$$

Again, $\hat{d}_\rho(f, g) = d_\infty(\text{epi } f, \text{epi } g)$ for $\rho = \infty$. Figure 10 illustrates $\hat{d}_\rho$, which involves “padding” the first epi-graph so that it includes the second one, and vice versa. In this figure, $\text{epi } g \cap \rho\mathbb{B}$ contains the single point $(0, \rho)$, which is clearly contained in $\text{epi } f$, and $\text{epi } g$ only needs to be padded with γ to ensure that $\text{epi } f \cap \rho\mathbb{B} \subset \text{epi } g + \gamma\mathbb{B}$. Consequently, $\hat{d}_\rho(f, g) = \gamma$.

Neither d_ρ nor $\hat{d}_\rho$ is a metric on $\text{lsc-fcns}(\mathbb{R}^n)$. The ρ -epi-distance d_ρ is a pseudometric, but $\hat{d}_\rho$ is not. Obviously, if $\rho\mathbb{B}$ is “small,” then $d_\rho(f, g)$ may very well be zero without f and g being equal. It is apparent that all values of ρ need to come into play. This is accomplished in the *epi-distance*

$$d(f, g) := \int_0^\infty d_\rho(f, g) e^{-\rho} d\rho,$$

which then leads to the following fact:

$$d \text{ is a metric on the space } \text{lsc-fcns}(\mathbb{R}^n).$$

We say that functions $f^\nu \in \text{lsc-fcns}(\mathbb{R}^n)$ *epi-converges* to a function $f \in \text{lsc-fcns}(\mathbb{R}^n)$ if $d(f^\nu, f) \rightarrow 0$. Epi-convergence is also characterized by d_ρ and $\hat{d}_\rho$:

For functions $f^\nu, f \in \text{lsc-fcns}(\mathbb{R}^n)$ and $\hat{\rho} \geq 0$, the following is equivalent:

$$\begin{aligned} d(f^\nu, f) &\rightarrow 0, \\ d_\rho(f^\nu, f) &\rightarrow 0 \quad \text{for } \rho \geq \hat{\rho}, \\ \hat{d}_\rho(f^\nu, f) &\rightarrow 0 \quad \text{for } \rho \geq \hat{\rho}. \end{aligned}$$

The epi-distance induces the *epi-topology* (in topology circles referred to as the Attouch–Wets topology) on $\text{lsc-fcns}(\mathbb{R}^n)$. In fact, by Theorem 7.58 of Rockafellar and Wets [18], $(\text{lsc-fcns}_{\neq\infty}(\mathbb{R}^n), \mathcal{d})$ is a complete metric space. It is also separable and therefore a Polish space, with, in fact, epi-splines given in terms of polynomials with rational coefficients furnishing a countable dense subset (Royset and Wets [19]). However, $\text{lsc-fcns}(\mathbb{R}^n)$ is not a vector space⁴ because $-f \notin \text{lsc-fcns}(\mathbb{R}^n)$ for $f \in \text{lsc-fcns}(\mathbb{R}^n)$ given by $f(x) = 0$ if $x \in [0, \infty)^n$ and $f(x) = 1$ otherwise. We note, however, that it is a cone, i.e., $\lambda f \in \text{lsc-fcns}(\mathbb{R}^n)$ for $\lambda \geq 0$, and $f \in \text{lsc-fcns}(\mathbb{R}^n)$. Moreover, $\text{lsc-fcns}(\mathbb{R}^n) \cup \{\infty\}$ is convex, where we have included the function that is infinity everywhere.

A parallel, slightly more general, development focusing on nonempty closed sets in $\mathbb{R}^n$, instead of epi-graphs of lsc functions, is also possible and leads to *set convergence* in the sense of Painlevé and Kuratowski. Such convergence is characterized by $\mathcal{d}$, $\mathcal{d}_\rho$, and $\hat{\mathcal{d}}_\rho$, similar to the above characterization of epi-convergence, with the slight change in definition of these distances consisting of replacing epi-graphs by closed sets; see Rockafellar and Wets [18, Chapter 4] and Beer [6] for further details about topologies on closed sets.

Connections between the various distances are provided by Rockafellar and Wets [18, Exercise 7.60]. For example, for functions f, g on $\mathbb{R}^n$ not identical to ∞ ,

- (i) $\hat{\mathcal{d}}_\rho(f, g) \leq \mathcal{d}_\rho(f, g) \leq \hat{\mathcal{d}}_{\rho'}(f, g)$, when $\rho' \geq 2\rho + \max\{d(0, \text{epi } f), d(0, \text{epi } g)\}$;
- (ii) $\mathcal{d}(f, g) \geq (1 - e^{-\rho})|d(0, \text{epi } f) - d(0, \text{epi } g)| + e^{-\rho}\mathcal{d}_\rho(f, g)$;
- (iii) $\mathcal{d}(f, g) \leq (1 - e^{-\rho})\mathcal{d}_\rho(f, g) + e^{-\rho}(\max\{d(0, \text{epi } f), d(0, \text{epi } g)\} + \rho + 1)$;
- (iv) $\mathcal{d}_\rho(f, g) = \hat{\mathcal{d}}_\rho(f, g)$ for any $\rho \geq 0$ if f, g convex with $f(0) \leq 0$ and $g(0) \leq 0$.

As is clear from these results, the distances are tied to the origin of $\mathbb{R}^{n+1}$. Although formulae for other “anchor” points are also possible, the simplest way of utilizing these estimates is to translate the functions f and g favorably. Item (iv) highlights the possibilities in that direction in the case of convex functions.

It is apparent that

$$\mathcal{d}_\rho(f, g) \leq \|f - g\|_\infty := \sup_{x \in \mathbb{R}^n} |f(x) - g(x)|,$$

and since $\int_0^\infty e^{-\rho} d\rho = 1$, also that

$$\mathcal{d}(f, g) \leq \|f - g\|_\infty.$$

Of course, with the possibilities of the function values $-\infty$ and ∞ , the right-hand sides may easily be ∞ , making reliance on $\|\cdot\|_\infty$ impractical. Even for finite-valued functions, it is clear that convergence in the epi-distance does not imply convergence in $\|\cdot\|_\infty$, as the following simple example shows. Suppose that $f^\nu(x) = 0$ if $x \leq 0$, and $f^\nu(x) = x/\nu$ otherwise. We also define $f^0(x) = 0$ for all $x \in \mathbb{R}$; see Figure 11. Clearly, $\|f^\nu - f^0\|_\infty = \infty$ for all ν . However, for any $\rho \geq 0$, $\mathcal{d}_\rho(f^\nu, f^0) \leq \rho/\nu$, and therefore $\mathcal{d}(f^\nu, f^0) \leq (1/\nu) \int_0^\infty \rho e^{-\rho} d\rho = 1/\nu$. Consequently, $\mathcal{d}(f^\nu, f^0) \rightarrow 0$, but $\|f^\nu - f^0\|_\infty \not\rightarrow 0$.

Pointwise convergence of a sequence of functions is also not equivalent to epi-convergence even if both pointwise and epigraphical limits exist, as the following simple example shows. Suppose that $f^\nu(x) = 0$ if $x = 1/\nu$, and $f^\nu(x) = 1$ otherwise. Clearly, the pointwise limit of f^ν , i.e., the function defined by $\lim_\nu f^\nu(x)$ for every x , is the function that is unity everywhere. In contrast, f^ν epi-converges to the function f^0 with $f^0(x) = 0$ for $x = 0$ and $f^0(x) = 1$ otherwise, i.e., $\mathcal{d}(f^\nu, f^0) \rightarrow 0$. In fact, $\mathcal{d}_\rho(f^\nu, f^0) = \hat{\mathcal{d}}_\rho(f^\nu, f^0) = 1/\nu$; see Figure 12.

⁴ We define the addition of functions and multiplication with a scalar in the usual “pointwise” manner. To handle extended real values, we adopt the conventions that $\infty + a = \infty$ and $-\infty + a = -\infty$ for $a \in \mathbb{R}$, $\infty + \infty = \infty + (-\infty) = -\infty + \infty = \infty$, $\lambda \cdot \infty = -\infty$ for $\lambda < 0$, $0 \cdot \infty = 0$, and $\lambda \cdot \infty = \infty$ for $\lambda > 0$, and similarly for $\lambda \cdot (-\infty)$.

FIGURE 11. Illustration of epi-convergence without uniform convergence.

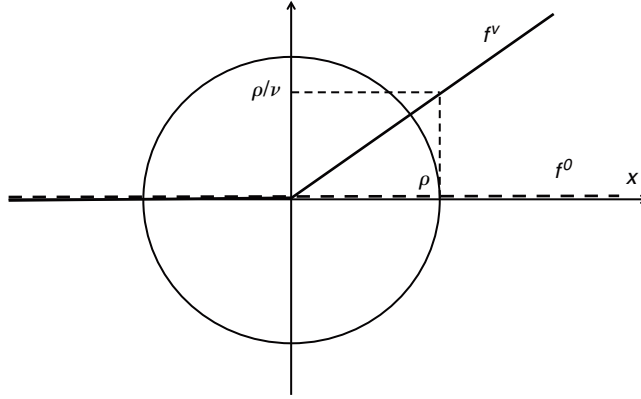
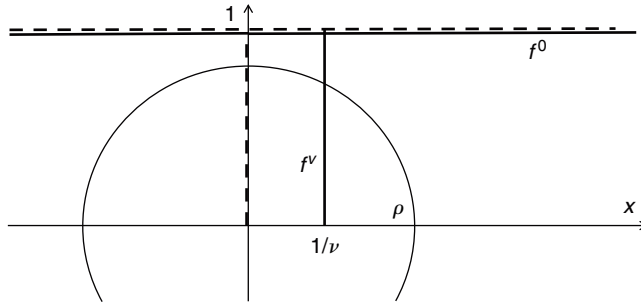


FIGURE 12. Illustration of epi-convergence and pointwise convergence.



Specific estimates for epi-splines are provided in Royset and Wets [19]. For $s, s' \in \text{e-spl}_n^p(\mathcal{R})$, with $\mathcal{R} = \{R_k\}_{k=1}^N$, one has, for any $\rho \geq 0$,

- (i) $d(s, s') \leq \max_{k=1, \dots, N} \sup_{x \in R_k} |s(x) - s'(x)|;$
- (ii) $d_\rho(s, s') \leq \max_{k=1, \dots, N} \sup_{x \in R_k} |s(x) - s'(x)|;$
- (iii) $\hat{d}_\rho(s, s') \leq \max_{k=1, \dots, N} \hat{d}_\rho(s_k, s'_k).$

These results provide a quantification of epi-convergence of functions in $\text{lsc-fcns}(\mathbb{R}^n)$. However, we need to go one step further and define epi-convergence of *optimization problems* defined over such functions.

We next provide a definition of epi-convergence of the approximate problems (P^ν) to the actual problem (P) . In view of the above results, these are problems defined on the metric space $(\text{lsc-fcns}(\mathbb{R}^n), d)$, and we therefore follow Attouch and Wets [2]:

A sequence $\{(P^\nu)\}_{\nu \in N}$ epi-converges to (P) if

- (i) for every sequence $\{s^\nu\}_{\nu \in K}$, with K an infinite subsequence of the positive integers, $s^\nu \in F^\nu \cap \mathcal{S}^\nu$, and $d(s^\nu, f) \rightarrow 0$, we have that $f \in F$ and $\liminf_\nu \psi^\nu(s^\nu) \geq \psi(f)$;
- (ii) for every $f \in F$, there exists a sequence $\{s^\nu\}_{\nu=1}^\infty$, with $s^\nu \in F^\nu \cap \mathcal{S}^\nu$, such that $d(s^\nu, f) \rightarrow 0$ and $\limsup_\nu \psi^\nu(s^\nu) \leq \psi(f)$.

The main consequence of epi-convergence in this context is the fact that it implies that solutions of the approximate problem tend to those of the actual problem. Let V and V^ν be the minimum values of (P) and (P^ν) , respectively. Specifically, by Attouch and Wets [2, Theorem 2.5],

if (P^ν) epi-converges to (P) , s^k minimizes (P^{ν_k}) , and $d(s^k, f) \rightarrow 0$, then f is a minimizer of (P) and $\lim_k V^{\nu_k} = V$.

This result therefore provides a direct path to an answer to the earlier question whether a solution of an approximate problem (P^ν) would be an approximate solution of (P) : one needs to ensure epi-convergence of (P^ν) to (P) .

5. Theory

A series of results are available about the relationship between (P^ν) and (P) and their solutions, as well as between epi-splines and lsc functions. Results in the context of probability density estimation are established in Royset and Wets [21]. We refer to Royset and Wets [20] for other cases with one-dimensional functions and Royset and Wets [19] for higher dimensions. Here, we give three central results dealing with epi-convergence of (P^ν) to (P) , the ability of epi-splines to approximate to an arbitrary level of accuracy any lsc function, and the rate of convergence of corresponding solutions.

Epi-convergence. A sufficient condition for epi-convergence of (P^ν) to the actual problem (P) relies on continuous convergence of the approximate criterion ψ^ν to the actual criterion ψ . Often, ψ^ν is simply identical to ψ for all ν . For the general case, we recall that ψ^ν converges continuously to ψ relative to $\mathcal{F}$ if for every $f \in \mathcal{F}$ and sequence $\{s^\nu\}_{\nu=1}^\infty$, with $s^\nu \in \mathcal{F}$ and $d(s^\nu, f) \rightarrow 0$, $\psi^\nu(s^\nu) \rightarrow \psi(f)$. We use the notation $\text{int } F$ to denote the interior of $F \subset \mathcal{F}$ (with respect to the epi-topology). We establish in Royset and Wets [19] that if

- (i) ψ^ν converges continuously to ψ relative to $\mathcal{F}$,
- (ii) $\bigcup_{\nu=1}^\infty \mathcal{S}^\nu$ is dense in $\mathcal{F}$, and
- (iii) F^ν set-converge to $F = \text{cl}(\text{int } F)$,

then (P^ν) epi-converges to (P) .

The first condition implies that ψ is continuous with respect to the epi-topology. The second condition relates to the ability of epi-splines to approximate arbitrary lsc functions, which we address in the next paragraph. The last part of the third condition is a constraint qualification that avoids “isolated” feasible points that cannot easily be approximated. Numerous simplification occurs if $\mathcal{F} \subset \text{e-spl}_n^p(\mathcal{R})$ for some partition $\mathcal{R}$ and order p , which implies that both (P^ν) and (P) are finite-dimensional, as discussed at the end of this section.

Approximation by epi-splines. To allow for approximation of arbitrary lsc functions, we consider a sequence of partitions that are gradually refined. Specifically, we say that a sequence $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$ of partitions of $\mathbb{R}^n$, with $\mathcal{R}^\nu = \{R_k^\nu\}_{k=1}^{N^\nu}$, is an *infinite refinement* if

$$\text{for every } x \in \mathbb{R}^n \text{ and } \varepsilon > 0 \text{ there exists a positive integer } \bar{\nu} \text{ such that} \\ R_k^\nu \subset \mathbb{B}(x, \varepsilon) \text{ for every } \nu \geq \bar{\nu} \text{ and } k \text{ satisfying } x \in \text{cl } R_k^\nu.$$

A simple example of an infinite refinement on $\mathbb{R}$ is to take $N^\nu = 2\nu + 2$, $R_1^\nu = (-\infty, -\sqrt{\nu})$, $R_k^\nu = ((k - \nu - 2)/\sqrt{\nu}, (k - \nu - 1)/\sqrt{\nu})$ for $k = 2, 3, \dots, 2\nu + 1$, and $R_{2\nu+2}^\nu = (\sqrt{\nu}, \infty)$. Then $\bar{\nu} > \max\{x^2, \varepsilon^{-2}\}$ satisfies the above condition. Obviously, much flexibility exists in constructing such infinite refinements.

A main result in Royset and Wets [19] is then that

for any nonnegative integer p and $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$ an infinite refinement on $\mathbb{R}^n$,

$$\bigcup_{\nu=1}^\infty \text{e-spl}_n^p(\mathcal{R}^\nu) \text{ is dense under the epi-topology in } \text{lsc-fncs}(\mathbb{R}^n).$$

Consequently, epi-splines, even those of order 0, can approximate to an arbitrary level of accuracy any lsc function. The optimization over epi-splines in (P^ν) is therefore justified.

Rate of convergence. Solutions of (P^ν) tend to those of (P) at rates that depend on several factors, and a full analysis along the lines of Attouch and Wets [4] and references therein is beyond the scope of this tutorial. Here, we consider the case in which $\mathcal{F} \subset \text{e-spl}_n^p(\mathcal{R})$ for some partition $\mathcal{R} = \{R_k\}_{k=1}^N$ and order p ; i.e., the actual problem (P) is finite dimensional. Of course, in view of the flexibility of epi-splines, there may be little to lose from a practical perspective to adopt this simplification from the outset. In view of the finite number of

parameters describing an epi-spline, (P) corresponds then to a parametric optimization problem

$$\text{Parametric actual problem } (\bar{P}): \min \bar{\psi}(r) \quad \text{such that } r \in R \subset \mathbb{R}^{n_e},$$

where R is a feasible region corresponding to the constraint set F , $\bar{\psi}$ is a criterion function expressed as a function of the parameters r , and n_e is defined in (1). If every $s \in \text{e-spl}_n^p(\mathcal{R})$ is of the form $s = \langle c(\cdot), r \rangle$, with $r \in \mathbb{R}^{n_e}$ and $c = (c_1, c_2, \dots, c_{n_e})$ a set of basis functions⁵ for the corresponding polynomials, then $\bar{\psi}(r) = \psi(\langle c(\cdot), r \rangle)$ and $R = \{r \in \mathbb{R}^{n_e} \mid \langle c(\cdot), r \rangle \in F\}$.

Similarly, a parametric optimization problem corresponding to (P^ν) takes the following form:

$$\text{Parametric approximate problem } (\bar{P}^\nu): \min \bar{\psi}^\nu(r) \quad \text{such that } r \in R^\nu,$$

where $R^\nu = \{r \in \mathbb{R}^{n_e} \mid \langle c(\cdot), r \rangle \in F^\nu \cap \mathcal{S}^\nu\}$ and $\bar{\psi}^\nu(r) = \psi^\nu(\langle c(\cdot), r \rangle)$. We quantify next the relationship between solutions of $(\bar{P}^\nu)$ and $(\bar{P})$.

We focus here on *near-optimal* solutions of $(\bar{P}^\nu)$ that are indeed the solutions provided by numerical solvers. A benefit of such an approach is that we avoid quantifying the “conditioning” of the problems’ near-optimal solutions; see Rockafellar and Wets [18, §7.J] for results in that direction. We therefore define the ε -optimal solutions for $(\bar{P}^\nu)$, which for $\varepsilon \geq 0$ are given by

$$\bar{R}_\varepsilon^\nu = \{r \in R^\nu \mid \bar{\psi}^\nu(r) \leq \bar{V}^\nu + \varepsilon\},$$

with $\bar{V}^\nu$ being the optimal value of $(\bar{P}^\nu)$. We also say that a function φ is Lipschitz on $\rho\mathbb{B}$ with constant κ if

$$|\varphi(r) - \varphi(r')| \leq \kappa \|r - r'\| \quad \text{for all } r, r' \in \rho\mathbb{B}.$$

A quantitative approximation result can then be deduced from Theorem 7.69 and Example 7.62 in Rockafellar and Wets [18]:

Suppose that $\bar{\psi}$, $\bar{\psi}^\nu$ are finite convex functions, R, R^ν are closed convex sets, and $\rho_0 \in (0, \infty)$ is such that there exists $r, r^\nu \in \rho_0\mathbb{B}$ optimal in $(\bar{P})$ and $(\bar{P}^\nu)$, respectively, and $\bar{\psi}(r), \bar{\psi}^\nu(r^\nu) \geq -\rho_0$. Let $\rho > \rho_0$, $\rho' > \hat{d}_\rho(R, R^\nu) + \rho$, and assume that $\bar{\psi}$, $\bar{\psi}^\nu$ are Lipschitz on $\rho'\mathbb{B}$ with constant κ . Then, for any $\varepsilon > 0$ and r^0 optimal in $(\bar{P})$,

$$d(r^0, \bar{R}_\varepsilon^\nu) \leq \left(1 + \frac{4\rho}{\varepsilon}\right) \left[\max_{r \in \rho\mathbb{B}} \{\bar{\psi}(r) - \bar{\psi}^\nu(r)\} + (\kappa + 1)\hat{d}_\rho(R, R^\nu)\right].$$

We observe that an error between an optimal solution of $(\bar{P})$ and $\bar{R}_\varepsilon^\nu$ originates from two sources: the differences between the criterion functions $\bar{\psi}$ and $\bar{\psi}^\nu$ and between the restrictions imposed through R and R^ν measured in terms of the $\hat{d}_\rho$ -distance, now defined for arbitrary sets and not only epi-graphs.

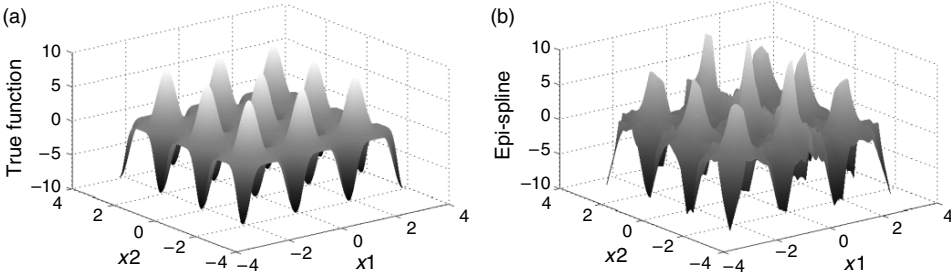
6. Examples

We provide numerical results for three gradually more complex examples. The first example estimates a common test function from image reconstruction. The second example estimates the probability density of simulation output. The third example forecasts electricity (load) demand on the basis of weather predictions.

Image reconstruction. We consider the standard test function $f(x) = (\cos(\pi x_1) + \cos(\pi x_2))^3$ defined on $[-3, 3]^2$; see Figure 13(a). We rely on continuous epi-splines and a partition of $[-3, 3]^2$ consisting of 20 by 20 squares so that $N = 400$. Based on 900 uniformly distributed data points and the max absolute error criterion, we obtain the epi-spline fit of Figure 13(b). Maximum and average errors across the data points are 0.415 and 0.313, respectively. Mean square and mean absolute error criteria give similar results. A relaxation of the continuity requirement results in essentially perfect interpolation at the expense of a more “rugged” fit.

⁵ There are several possible basis functions with $(1, x, x^2, \dots, x^p)$ for each interval (m_{k-1}, m_k) furnishing one example on $\mathbb{R}$; see (2).

FIGURE 13. Image reconstruction example: Actual function (a) and epi-spline (b).



Simulation output estimation. Estimates of the probability density function of the output of simulation model provide a comprehensive picture of the performance of the system modelled. Although kernel methods and other traditional methods of nonparametric statistics compute density estimates, they rely heavily on large sample sizes and only in part can account for external information, a prevalent factor in practical simulation studies. As laid out in Royset and Wets [21], probability density estimation using epi-splines offers the possibility to include an arbitrary collection of external information and thereby achieve high-quality estimates even for small sample sizes.

For a sample $X^1, X^2, \dots, X^\nu$ that is independently and identically distributed according to an unknown density, it is well known (see, for example, Royset and Wets [21]) that a constrained maximum likelihood estimator of the density is an optimal solution of

$$\max \frac{1}{\nu} \prod_{i=1}^{\nu} h(X^i)^{1/\nu} \quad \text{such that } h \geq 0, \quad \int h(x) dx = 1, \quad h \in \mathcal{H},$$

where $\mathcal{H}$ is an appropriately selected space of functions on $\mathbb{R}^n$. Passing through the transformation $h = e^{-f}$, we arrive after equivalently maximizing the logarithm of the objective function of the problem

$$\min \frac{1}{\nu} \sum_{i=1}^{\nu} f(X^i) \quad \text{such that } \int e^{-f(x)} dx = 1, \quad f \in \mathcal{F},$$

where, as above, we let $\mathcal{F} \subseteq \text{lsc-fncs}(\mathbb{R}^n)$. Since this problem is infinite-dimensional, we formulate the approximate problem

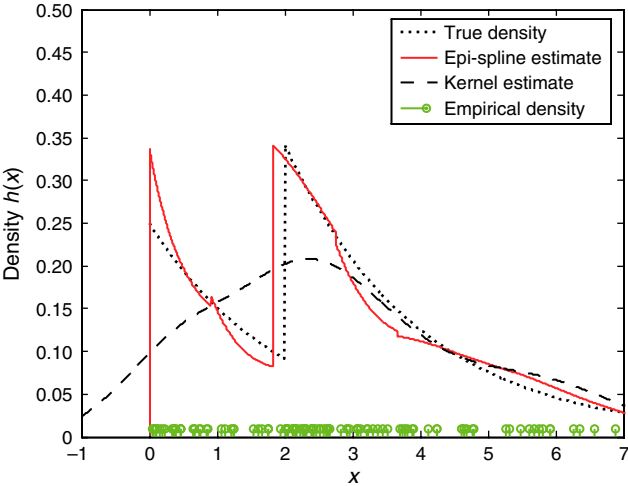
$$\min \frac{1}{\nu} \sum_{i=1}^{\nu} s(X^i) \quad \text{such that } s \in F^\nu \cap \mathcal{S}^\nu$$

over epi-splines, where $\mathcal{S}^\nu = \text{e-spl}_n^p(\mathcal{R}) \cap \mathcal{F}$, as above, and F^ν is a subset of functions satisfying $\int e^{-s(x)} dx = 1$ but also satisfying other restrictions derived from external information. A solution s of the approximate problem provides a density estimator through the composition e^{-s} , where we observe that the nonnegativity is automatically satisfied.

An example taken from Singham et al. [24] illustrates the possibilities. We consider an $M/M/1$ queue with arrival rate $\lambda = 1$ and service rate $\mu = 1.5$, but 50% of customers who enter the system are held at a separate station for two time units before entering the queue. We wish to estimate the density of the customer time-in-service (TIS). The actual density is an equal mixture of the density corresponding to an $M/M/1$ queue without the holding station, which is exponential, and the same density shifted to the right by two time units. The resulting density is discontinuous at two time units because half the customers will have two time units added to their time in system; see the dotted line in Figure 14.

We assume that this detailed information about the density for customer TIS is unavailable and that we must rely on 100 simulated customer TIS values as well as reasonable, external

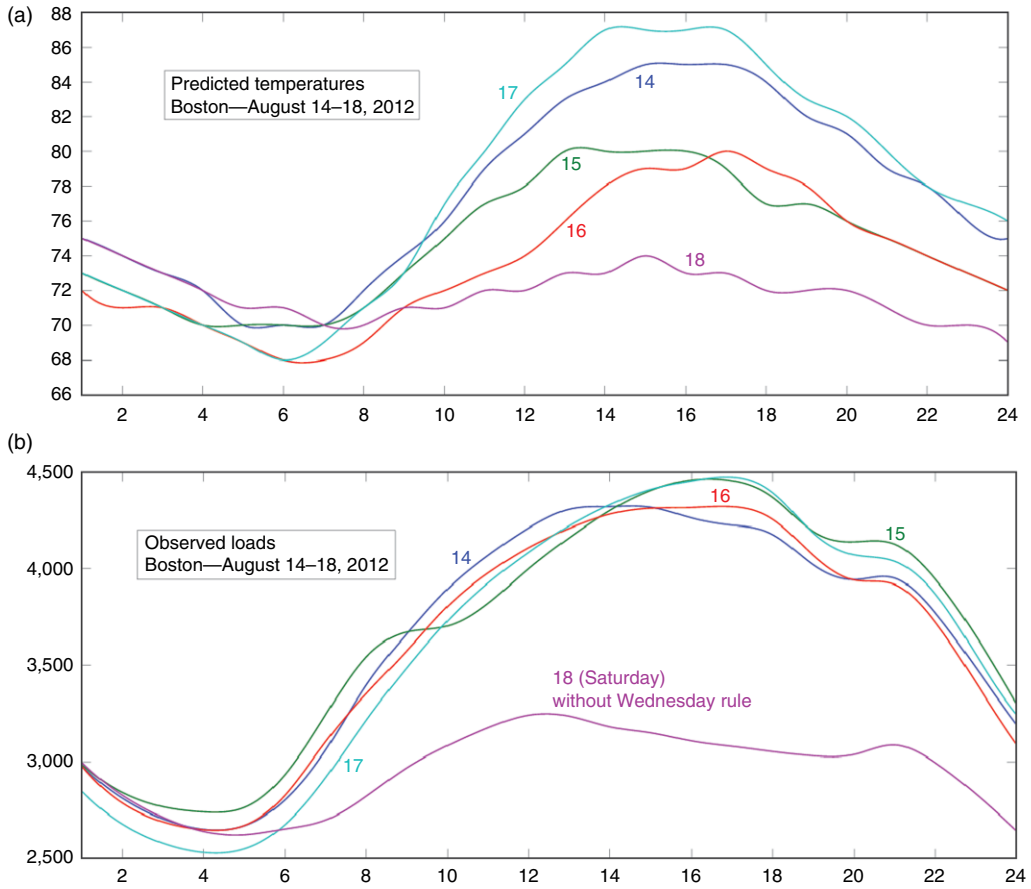
FIGURE 14. Density of queuing simulation.



information. Of course, in practice this is the common situation. We build a second-order epi-spline estimate by solving the above approximate problem, under the external information that the customer TIS is nonnegative, derivatives relative to density values are bounded (i.e., a bound on the pointwise Fisher information), second-order derivatives are bounded, and the upper tail is log concave. Using a mesh consisting of $N = 10$ intervals, we obtain the estimate given by the solid line in Figure 14. The picture is representative of hundreds of draws of 100 simulated customer TIS values. We also illustrate a standard Gaussian kernel estimate. The stems indicate the 100 simulated customer TIS values. The estimate based on epi-splines tracks the actual density closely and, in contrast to the kernel estimate, captures its essence. From this and similar examples in Singham et al. [24], we conclude that epi-spline technology provides a tool for analyzing simulation output under a variety of external information. In particular, flexibility of this kind is important when simulations are expensive and only a small sample can be made available. The burden then falls on the external sources to provide information that can sharpen performance estimates and support decisions about the underlying system.

Electricity load forecast. A significant portion of electricity supply in the United States is provided through “day-ahead” contracts, where producers agree to deliver certain quantities of electricity for the next day. A central component of a marketplace for such contracts is forecasting tools for the next day’s electricity load (demand). Such tools must necessarily rely on the information available at the time of forecasting, which typically includes weather forecasts for the next day (temperature, dew point, cloud cover, etc.) as well as historical information about electricity load on days that had similar weather forecasts. Figure 15(a) shows temperature forecasts for five days in Boston in 2012, with corresponding actually observed electricity loads on those days shown in Figure 15(b). Although, the historical information may stretch back decades, the useful information is much more limited because the data need to be carefully segmented to ensure that only “similar” days are included; see Feng et al. [13, 14]. For example, data for Mondays should not be mixed with those for Saturdays because the electricity load is fundamentally different on the weekend, and data for hot days should not be mixed with those of cool days. In the end, only about 15 to 25 days per year may remain on which the forecast must rely, making traditional techniques based on time-series or stochastic differential equations inaccurate. Here, we briefly describe the epi-spline-based construction of a stochastic process of the next day’s electricity load as laid out in Feng et al. [13, 14]; see also Rios et al. [16]. Epi-splines enter at three points. First, to help segment the data, probability density functions are estimated using an approach similar to that described for simulation output above. Second, a regression model builds an estimate

FIGURE 15. Temperature forecasts (a) and corresponding actual loads (b) for five days.



of the “expected” load for the full 24 hours of the next day. Third, conditional probability density estimates quantify the variability in the process. Here, we describe the two last steps only.

Suppose that there are J days of historical data on which to base the estimates. For each day $j = 1, 2, \dots, J$ and hour $h = 1, 2, \dots, 24$ of that day, we know the forecasted temperature t_h^j , the forecasted dew point d_h^j , and the actually observed load l_h^j . Additional weather statistics are also available but excluded here for simplicity of exposition. In any case, temperature and dew point forecasts appear to suffice for the estimation of summer days (Feng et al. [13, 14]). The regression model then takes the form

$$l_h^j = s^{\text{tmp}}(h)t_h^j + s^{\text{dpt}}(h)d_h^j + e_h^j,$$

where s^{tmp} and s^{dpt} are epi-splines defined on $[0, 24]$, and e_h^j is the error between the observed load l_h^j and the predicted load $s^{\text{tmp}}(h)t_h^j + s^{\text{dpt}}(h)d_h^j$. We note that here the “coefficients” in front of the explanatory variables *temperature* and *dew point* are functions of time. The regression problem is then to determine epi-splines s^{tmp} and s^{dpt} that

$$\text{minimize } \sum_{j=1}^J \sum_{h=1}^{24} |l_h^j - s^{\text{tmp}}(h)t_h^j - s^{\text{dpt}}(h)d_h^j|,$$

with the external information that the epi-splines are of second order, are continuously differentiable, and have bounded second-order derivatives. Using the representation of

polynomials in (2) and mesh points $m_k = k24/N$, $k = 1, 2, \dots, N$, partitioning $[0, 24]$ into N intervals, we obtain that

$$s^{\text{tmp}}(x) = a_0^{k, \text{tmp}} + a_1^{k, \text{tmp}}x + a_2^{k, \text{tmp}}x^2 \quad \text{for } x \in (m_{k-1}, m_k),$$

with continuity forcing the epispline on the mesh points to be defined by the values at the adjacent intervals; s^{dpt} is defined similarly in terms of parameters $a_0^{k, \text{dpt}}$, $a_1^{k, \text{dpt}}$, and $a_2^{k, \text{dpt}}$. Let $m_0 = 0$. The regression problem then takes the form of a linear program over those parameters as well as over the auxiliary parameters z_h^j :

$$\begin{aligned} \min \quad & \sum_{j=1}^J \sum_{h=1}^{24} z_h^j \quad \text{such that} \\ & l_h^j - (a_0^{k, \text{tmp}} + a_1^{k, \text{tmp}}h + a_2^{k, \text{tmp}}h^2)t_h^j - (a_0^{k, \text{dpt}} + a_1^{k, \text{dpt}}h + a_2^{k, \text{dpt}}h^2)d_h^j \leq z_h^j, \\ & \quad h \in (m_{k-1}, m_k] \quad k = 1, \dots, N, \quad j = 1, \dots, J; \end{aligned} \quad (3)$$

$$\begin{aligned} & -l_h^j + (a_0^{k, \text{tmp}} + a_1^{k, \text{tmp}}h + a_2^{k, \text{tmp}}h^2)t_h^j + (a_0^{k, \text{dpt}} + a_1^{k, \text{dpt}}h + a_2^{k, \text{dpt}}h^2)d_h^j \leq z_h^j, \\ & \quad h \in (m_{k-1}, m_k] \quad k = 1, \dots, N, \quad j = 1, \dots, J; \end{aligned} \quad (4)$$

$$\begin{aligned} & a_0^{k, \text{tmp}} + a_1^{k, \text{tmp}}m_k + a_2^{k, \text{tmp}}m_k^2 - (a_0^{k+1, \text{tmp}} + a_1^{k+1, \text{tmp}}m_k + a_2^{k+1, \text{tmp}}m_k^2) = 0, \\ & \quad k = 1, \dots, N-1; \end{aligned} \quad (5)$$

$$a_1^{k, \text{tmp}} + 2a_2^{k, \text{tmp}}m_k - (a_1^{k+1, \text{tmp}} + 2a_2^{k+1, \text{tmp}}m_k) = 0, \quad k = 1, \dots, N-1; \quad (6)$$

$$-\kappa \leq a_2^{k, \text{dpt}}, \quad a_2^{k, \text{tmp}} \leq \kappa, \quad k = 1, \dots, N. \quad (7)$$

The inequalities (3) and (4) linearize the criterion function, (5) and (6) ensure continuity and continuous differentiability requirements (see §3.3), and (7) bounds the magnitude of the second-order derivatives to κ . In Feng et al. [14], $N = 24$ and $\kappa = 500$. A slightly more complicated version with weighted errors is also possible.

The regression curve can be viewed as the drift of the stochastic process that forecasts the electricity load. The next step is to determine the volatility of the process. For a fixed hour h , we can interpret the minimized errors $\{e_h^j\}_{j=1}^J$ as the observations associated with the distribution of these errors. Using the epi-spline-based procedure for estimating probability densities described above, it is then possible to estimate this distribution. Since the number of observed errors will be limited, it becomes especially critical to include external information to ensure quality estimates. We can repeat this process for each hour h independently, but that would not take into account *conditioning*. Clearly, if the load at hour h is substantially higher than the load tentatively projected by the regression curve, one should take this into account when looking at the distribution of the errors at a later time $h + \Delta h$. This can be done systematically by restricting the samples of the error at time $h + \Delta h$ to those that come from (observed) load trajectories that at time h had similar deviations from the overall drift of the process, i.e., from the regression curve (for details, see Feng et al. [13, 14], Rios et al. [16]). Since this limits the number of data points on which a single density estimate must rely, the significance of being able to include external information in the epi-spline framework is further highlighted.

Extensive numerical results are provided by Feng et al. [13, 14] for case studies covering the New England region of the United States. Figure 16 from Feng et al. [13] illustrates two epi-spline-based estimates of the probability density function of errors at hours $h = 6$ and 17 on a particular day. For a day with high variability in electricity load taken from Feng et al. [14], Figure 17 gives an example of estimated loads in 2011 in Connecticut according to the regression model (line with circles), the associated uncertainty given in terms of alternative scenarios (lines with solid dots), and the actual load (line with squares). It is clear that

FIGURE 16. Probability densities of errors during two hours of the day.

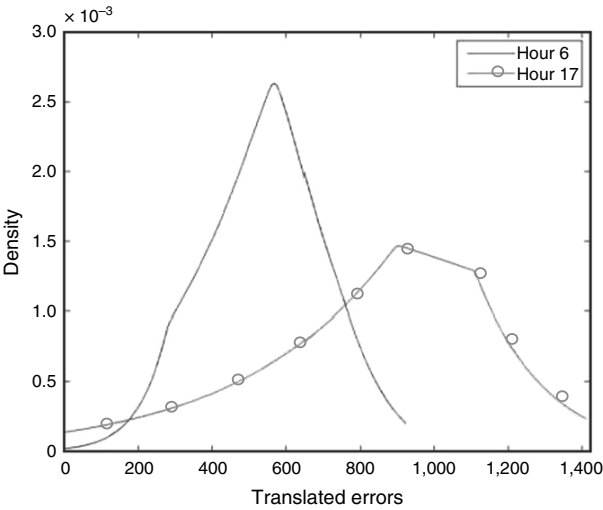
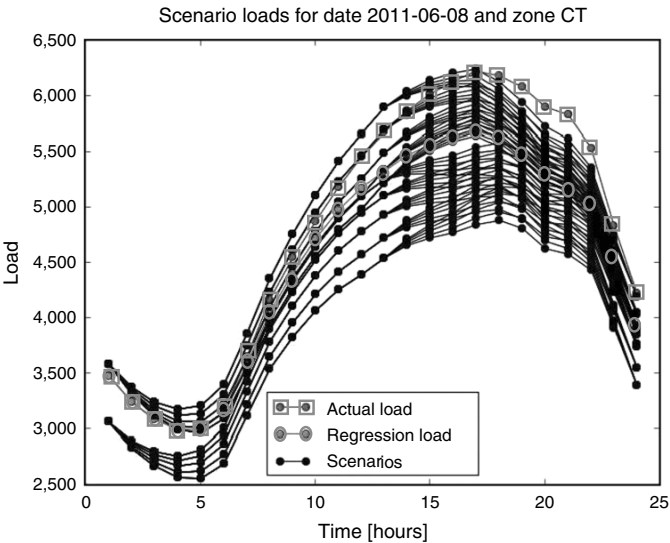


FIGURE 17. Regression load with estimated uncertainty and actual load.



the constructed stochastic process covers the actual load to a high degree. Table 1 taken from Feng et al. [14] shows aggregated forecasting error statistics in terms of mean average percentage forecasting errors from actual loads for different seasons and numbers of segments used in the procedure for segmenting the data into clusters of “similar” days. We find that, regardless of the season, the forecasting errors are small, especially under an intelligent segmentation of the data. The epi-spline-based forecasting method of Feng et al. [13, 14] is

TABLE 1. Mean average percent forecasting error in electric load forecast.

Season	Numbers of segments			
	1	3	5	7
Fall	5.45	4.66	4.2	3.99
Spring	3.1	2.88	2.67	2.73
Summer	10.25	4.82	4.14	4.19
Winter	5.25	3.32	3.29	3.47

currently being implemented by an independent software provider, with planned adoption by the New England Independent System Operator to support daily planning in 2014.

Acknowledgments

This material is based upon work supported in part by the United States Army Research Laboratory and the United States Army Research Office [Grants 00101-80683, W911NF-10-1-0246, and W911NF-12-1-0273], and, in part, was carried out while the first author was on a sabbatical leave at the University of California (UC) Department of Mathematics. The authors also thank B. Chern (Stanford University), S. Ryan (Iowa State University), D. Singham (Naval Postgraduate School), J.-P. Watson (Sandia Laboratory), and D. Woodruff (UC Davis) for providing information about epi-spline applications.

References

- [1] M. Atteia. Généralisation de la définition et des propriétés de “splines-fonctions.” *Comptes Rendus de l'Académie des Sciences de Paris* 260:3350–3553, 1965.
- [2] H. Attouch and R. J-B Wets. A convergence theory for saddle functions. *Transactions of the American Mathematical Society* 280(1):1–41, 1983.
- [3] H. Attouch and R. J-B Wets. Epigraphical analysis. *Analyse Non Linéaire*. Gauthier-Villars, Paris, 73–100, 1989.
- [4] H. Attouch and R. J-B Wets. Quantitative stability of variational systems: III. ε -approximate solutions. *Mathematical Programming* 61(1):197–214, 1993.
- [5] H. Attouch, R. Lucchetti, and R. J-B Wets. The topology of the ρ -Hausdorff distance. *Annali di Matematica pura ed applicata* CLX:303–320, 1991.
- [6] G. Beer. *Topologies on Closed and Closed Convex Sets*. Mathematics and Its Applications, Vol. 268, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992.
- [7] M. Casey and R. J-B Wets. Density estimation: Exploiting nondata information. Technical report, University of California, Davis, Davis, CA, 2010.
- [8] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Annals of Statistics* 41(5):2428–2461, 2013.
- [9] H. B. Curry and I. J. Schoenberg. On spline distributions and their limits: The Polya distribution functions. *Bulletin of the American Mathematical Society* 53:1114, 1947.
- [10] H. B. Curry and I. J. Schoenberg. On Polya frequency functions IV: The fundamental spline functions and their limits. *Journal d'Analyse Mathématique* 17:71–107, 1966.
- [11] C. de Boer. Best approximation properties of spline functions of odd degree. *Journal of Mathematical Mechanics* 12:747–749, 1963.
- [12] M. Egerstedt and C. Martin. *Control Theoretic Splines*. Princeton University Press, Princeton, NJ, 2010.
- [13] Y. Feng, D. Gade, S. M. Ryan, J.-P. Watson, R. J-B Wets, and D. L. Woodruff. A new approximation method for generating day-ahead load scenarios. *Proceedings of the IEEE Power and Energy Society General Meeting*, IEEE, Piscataway, NJ, 2013.
- [14] Y. Feng, I. Rios, S. M. Ryan, K. Spürkel, J.-P. Watson, R. J-B Wets, and D. L. Woodruff. Toward scalable stochastic unit commitment—Part I: Load scenario generation. Working paper, Iowa State University, Ames, IA, 2013.
- [15] J. C. Holladay. A smoothest curve approximation. *Mathematical Tables and Other Aids to Computation* 11:233–243, 1957.
- [16] I. Rios, R. J-B Wets, and D. L. Woodruff. Multi-period forecasting and scenarios generation with limited data. *Computational Management Science*. Working paper, University of California, Davis, Davis, CA, 2014.
- [17] R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* 18(1–2):33–53, 2013.
- [18] R. T. Rockafellar and R. J-B Wets. *Variational analysis*. Grundlehren der Mathematischen Wissenschaft, Vol. 317, 3rd printing, Springer, New York, 2009.

- [19] J. O. Royset and R. J-B Wets. Function identification and multi-dimensional epi-splines. Working paper, University of California, Davis, Davis, CA, 2014.
- [20] J. O. Royset and R. J-B Wets. Evolving function identification problems: Epi-splines as pliable approximations tools. Working paper, University of California, Davis, Davis, CA, 2014.
- [21] J. O. Royset and R. J-B Wets. Fusion of soft and hard information in nonparametric density estimation. Working paper, University of California, Davis, Davis, CA, 2014.
- [22] J. O. Royset, N. Sukumar, and R. J-B Wets. Uncertainty quantification using exponential epi-splines. *Proceedings of the International Conference on Structural Safety and Reliability*, 2013.
- [23] I. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics* 4:45–99, 112–141, 1946.
- [24] D. I. Singham, J. O. Royset, and R. J-B Wets. Density estimation of simulation output using exponential epi-splines. *Proceedings of the 2013 Winter Simulation Conference*, IEEE, Piscataway, NJ, 755–765, 2013.
- [25] R. Sood and R. J-B Wets. Information fusion. Accessed May 2014, <http://www.math.ucdavis.edu/~prop01>, 2011.
- [26] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
- [27] R. J-B Wets and S. Bianchi. Term and volatility structures. S. Zenios and W. Ziemba, eds. *Handbook of Asset and Liability Management*. North Holland, Amsterdam, 26–68, 2006.
- [28] R. J-B Wets and I. Rios. Modeling and estimating commodity prices: Copper prices. *Mathematics of Finance and Economics*. Working paper, University of California, Davis, Davis, CA, 2013.
- [29] R. J-B Wets, S. Bianchi, and L. Yang. Serious zero-curves. EpiRisk. Accessed May 2014, <http://epirisk.com/epicurvespub.php4>, 2002.