Astronomy
&
Astrophysics

Interpolation and smoothing

M. Lombardi[⋆]

Institut für Astrophysik und Extraterrestrische Forschung, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

Abstract. Smoothing is omnipresent in astronomy, because almost always measurements performed at discrete positions in the sky need to be interpolated into a smooth map for subsequent analysis. Still, the statistical properties of different interpolation techniques are very poorly known. In this paper, we consider the general problem of interpolating discrete data whose location measurements are distributed on the sky according to a known density distribution (with or without clustering). We derive expressions for the expectation value and for the covariance of the smoothed map for many interpolation techniques, and obtain a general method that can be used to obtain these quantities for any linear smoothing. Moreover, we show that few basic properties of smoothing procedures have important consequences on the statistical properties of the smoothed map. Our analysis allows one to obtain the statistical properties of an arbitrary interpolation procedure, and thus to optimally choose the technique that is most suitable for one's needs.

Key words. methods: statistical – methods: analytical – methods: data analysis – gravitational lensing

1. Introduction

A common problem in astronomy is the smoothing of irregularly sampled data. In general, this happens when one can make measurements on discrete points of a quantity that has some astrophysical relevance. In most cases, the "quantity" is a *continuous field*, i.e. a smooth function on the sky. In this case it is reasonable to try to reconstruct the field by interpolating the discrete measurements obtained.

Suppose, for example, that we are interested in measuring the column density of a nearby molecular cloud. A good estimate of the cloud column density can be obtained, for example, using the infrared color of background stars observed through the cloud (see, e.g., Lombardi & Alves 2001). This way, we can obtain reliable measurements of the "column density field" at the discrete points corresponding to the star positions. Finally, we interpolate the various measurements and obtain a smooth map. Similar situations are often encountered in different fields (e.g., weak gravitational lensing, peculiar velocity field of galaxies).

Interpolation and smoothing are ubiquitous in astronomy, and indeed many papers have been devoted to the study of the effects of interpolation in particular analyses (see, e.g. Rybicki & Press 1992). However, it is important to observe that many different approaches to the study of interpolation are possible, depending on the type of problem considered. In this paper, in particular, we will study the statistical properties of several interpolation techniques. Since the locations on the sky where measurements are performed cannot generally be chosen in Astronomy (cf. the example of reddening of stars above),

we will carry out an *ensemble average* over the measurement locations. More precisely, we will assume that the measurements are randomly distributed on the sky following a known scheme and we will carry out a statistical analysis on this sample. Note that there is a complete freedom on the spatial distribution of measurements on the sky, that can be correlated (for example, clustered) or can follow a non-uniform density $\rho(x)$. The ensemble average, which has already been carried out under simplifying hypotheses of uncorrelated points and uniform density in earlier papers (e.g., Lombardi & Bertin 1998; van Waerbeke 2000; Lombardi & Schneider 2001, hereafter Paper I; Lombardi et al. 2002; Lombardi & Schneider 2002, hereafter Paper II), let us to derive general results that generalize the particular configuration considered.

In previous works (see in particular Papers I and II) we have focused on a widely used smoothing method. Here, in contrast, we will keep the discussion much more general and consider some wide classes of interpolating techniques. This alternative approach has a number of advantages: (i) the results obtained are very general and can be applied to several smoothing techniques ("proof reusability"); (ii) the analytic discussion is kept at a very simple level; (iii) the results obtained are more general, since allow for a non-uniform density of locations and correlation on the locations; (iv) the properties of a smoothing technique can be predicted in advance without the need of long calculations. This last point, in our opinion, is particularly important for practical applications. In fact, with the aid of the results obtained in this paper, several smoothing techniques can be easily compared, which allows one to choose the interpolation method more suitable. We stress, however, that some specific results cannot be derived using the techniques described in

⋆ e-mail: lombardi@astro.uni-bonn.de

this article and need a more specific analysis as done in Papers I and II.

The paper is organized in two main parts. In the first part, Sect. 2, we classify the smoothing methods and we show the general results associated to each interpolation family. In the second part, Sect. 3, we illustrate the results obtained earlier by considering several common smoothing techniques. Finally, in Sect. 4, we briefly draw the conclusions of this analysis.

1.1. Notation

Let us suppose to have a set of pairs $\{(x_i, y_i)\}$, where $x_i \in X$, called "locations", belong to a real vector space X (typically, $X = \mathbb{R}^n$, with $n = 1, 2$, or 3), and $y_i \in Y$, called "values", belong to a field Y (in this paper, we will assume for simplicity $Y = \mathbb{R}$). The spatial interpolation problem consists in finding a way to obtain approximated values for a generic $x \in X$. Formally, an interpolation procedure is a function S that maps a point $x \in X$ and an *unordered* set of couples $(x_i, y_i) \in X \times Y$ into a value $y \in Y$:

$$y = S(x; \{(x_i, y_i)\}). \tag{1}$$

Note that the number N of couples (x_i, y_i) is not fixed a priori (in particular N could vanish). In the following we will use as synonymous interpolation procedure, smoothing technique, interpolator.

As an example, let us consider a simple interpolator where the smoothed value at x is obtained as a weighted sum of the values $\{y_i\}$. More precisely, we define S as

$$S(x; \{(x_i, y_i)\}) = \frac{\sum_{j=1}^{N} y_j / |x - x_j|^\alpha}{\sum_{j=1}^{N} 1 / |x - x_j|^\alpha}, \tag{2}$$

where α is a fixed real number. This interpolation technique will be often used in this paper to illustrate with an example some general, abstract results.

In this paper we will study some statistical properties of various interpolation procedures assuming that the locations $\{x_i\}$ are random variables distributed with spatial density $\rho(x)$ (with or without clustering), and that the values $\{y_i\}$ are associated to the locations through the relation

$$y_i = f(x_i) + \epsilon_i. \tag{3}$$

Here $f : X \to Y$ is a known function and $\{\epsilon_i\}$, representing measurement errors, are random variables with vanishing mean and covariance matrix (taking $\{\epsilon_i\}$ as a multidimensional random variable) proportional to the identity:

$$\langle \epsilon_i \rangle = 0, \qquad \langle \epsilon_i \epsilon_j \rangle = \delta_{ij} \sigma^2. \tag{4}$$

Note that the second relation of Eq. (4) states the so-called "statistical orthogonality" of the measurement errors; this property is trivially satisfied if $\{\epsilon_i\}$ are independent random variables with fixed variance σ^2 (this is a good approximation for many astronomical observations). The subject of our study will be the average value $\langle \tilde{f}(x) \rangle$ of $\tilde{f}(x)$, where $\tilde{f}(x)$ is the interpolated value of f at x:

$$\tilde{f}(x) = S(x; \{(x_i, y_i)\}) = S(x; \{(x_i, f(x_i) + \epsilon_i)\}). \tag{5}$$

We will also investigate the covariance (or two-point correlation) of the map $\tilde{f}(x)$, defined as

$$\begin{aligned}
\mathrm{Cov}(x, x'; \tilde{f}) &= \left\langle [\tilde{f}(x) - \langle \tilde{f}(x) \rangle][\tilde{f}(x') - \langle \tilde{f}(x') \rangle] \right\rangle \\
&= \langle \tilde{f}(x)\tilde{f}(x') \rangle - \langle \tilde{f}(x) \rangle \langle \tilde{f}(x') \rangle.
\end{aligned} \tag{6}$$

A word of explanation is needed regarding averages. Averages of $\tilde{f}(x)$ and $\tilde{f}(x)\tilde{f}(x')$ are carried out both with respect to $\{\epsilon_i\}$ and to $\{x_i\}$. As we have anticipated above, averages with respect to $\{x_i\}$ are carried out assuming that the locations are random variables distributed with density $\rho(x)$ over the set X. For example, if we assume that the locations are uncorrelated and that the density is constant over the field (this is usually referred to as a homogeneous Poisson process or as compete spatial randomness), then the number N of object locations inside a field $A \subset X$ of finite area $\mu(A)$ follows a Poisson distribution with average $\rho\mu(A)$:

$$p_N(N) = \mathrm{e}^{-\rho\mu(A)} \frac{\left[\rho\mu(A) \right]^N}{N!}. \tag{7}$$

Under the same hypotheses, the N locations $\{x_i\}$ are, then, uniformly distributed inside A. In more general cases (in particular, in case of correlation), it can be non-trivial to specify exactly the probability distribution of points. However, fortunately for the following discussion we need only the density $\rho(x)$ and an algorithm to randomly generate the locations (as discussed by Scott et al. 1954; see also, e.g., the hierarchical model described by Soneira & Peebles 1978).

The probability distribution for the other random variables considered here, namely $\{\epsilon_i\}$, need not to be fully specified. Actually, for our purposes, we just need to specify that these variables, representing measurement errors, have vanishing mean, fixed variance, are orthogonal (i.e., satisfy Eq. (4)), and are independent of the locations $\{x_i\}$. In the following calculations, we will carry out first the average over the measurement errors $\{\epsilon_i\}$, and then the one over the locations $\{x_i\}$.

We finally note that in this study we allow for cases where the smoothing procedure S cannot be defined. For example, if the density ρ is small, we might end up with a configuration without locations x_i close enough to x. Some interpolation procedures then might not be applied. In the following, we adopt the convention of discarding in the ensemble average the configurations $\{x_i\}$ for which the interpolation procedure (1) is not *locally* defined. For example, when evaluating the average $\langle \tilde{f}(x) \rangle$, we discard configurations that make S not defined at x; similarly, for expressions such as $\langle \tilde{f}(x)\tilde{f}(x') \rangle$ we discard configurations for which S is not defined at x or at x'. Note that we assume that the applicability of the smoothing technique depends only on the locations $\{x_i\}$ and *not* on the values $\{y_i\}$. We call $P_0(x)$ the probability of having a configuration $\{x_i\}$ for which the estimator is not defined at x. For example, some smoothing techniques (see Sect. 2.6) are not defined if there are no locations inside a given subset π; in this case we find (in case of vanishing correlation, i.e. for a Poisson distribution of locations)

$$P_0(x) = \exp\left(-\int_\pi \rho(x)\,\mathrm{d}x \right). \tag{8}$$

Similarly, we define $P_0(x, x')$ as the probability that the smoothing is not defined either at x or at x' (or at both points). If, again, the smoothing technique is defined at x only if there is at least a location inside a given subset π, and is defined at x' if there is a location inside the subset π', then

$$
\begin{aligned}
P_0(x, x') &= \exp\left(-\int_{\pi \cap \pi'} \rho(x)\,dx\right) \\
&\quad \times \left[\exp\left(-\int_{\pi \setminus \pi'} \rho(x)\,dx\right) + \exp\left(-\int_{\pi' \setminus \pi} \rho(x)\,dx\right)\right. \\
&\quad \left. - \exp\left(-\int_{(\pi \setminus \pi') \cup (\pi' \setminus \pi)} \rho(x)\,dx\right)\right].
\end{aligned}
\tag{9}
$$

Note that $P_0(x, x) = P_0(x)$ and that $P_0(x)P_0(x') \le P_0(x, x') \le P_0(x) + P_0(x')$. If an estimator is always defined, as for the interpolator (2), we just set $P_0(x) = P_0(x, x') = 0$.

2. Classification of interpolation procedures

Very little could be said at this stage about the average of $\tilde{f}(x)$ or its covariance because the problem in the formulation of Sect. 1.1 is by far too general. However, as we will see below, additional hypotheses allow us to obtain various results. Some of these hypotheses are quite obvious, in the sense that we could hardly imagine a sensible interpolation procedure that does not satisfy them. For example, if all values are the same, say $y_i = 1$, we expect to have always $y = 1$ in Eq. (1) (this property is discussed in Sect. 2.2). Still this "obvious" hypothesis allows us to obtain some non-trivial results.

In this section we will consider some criteria used to classify interpolation procedures, with the aim of introducing a standard terminology and, at the same time, of deriving a number of useful results.

2.1. Linearity

Some of the most interesting interpolation procedures are linear functions of the data values $\{y_i\}$. We stress that the linearity is on the *values* and not on the locations $\{x_i\}$. Formally, the linearity is expressed by the relation

$$
S(x; \{(x_i, \alpha y_i + \beta y_i')\}) = \alpha S(x; \{(x_i, y_i)\}) + \beta S(x; \{(x_i, y_i')\}), \tag{10}
$$

where α and β are arbitrary real numbers. Given a location configuration $\{x_i\}$, we define the jth *weight* of S as

$$
w_j(x; \{x_i\}) = S(x; \{(x_i, \delta_{ij})\}). \tag{11}
$$

Here δ_{ij} is the Kronecker symbol. Using Eq. (11), we can write any linear interpolation procedure in the form

$$
S(x; \{(x_i, y_i)\}) = \sum_{j=1}^{N} w_j(x; \{x_i\}) y_j. \tag{12}
$$

For example, the smoothing technique (2) is manifestly a linear function of the values $\{y_i\}$, and thus is a linear interpolator. Actually, it is already written in the form (12) with

$$
w_j(x; \{x_i\}) = \frac{1/|x - x_j|^\alpha}{\sum_{i=1}^{N} 1/|x - x_i|^\alpha}. \tag{13}
$$

Since the discussion of linear smoothing techniques is quite lengthy, we split it into two subsections, corresponding to the study of the bias and the covariance of the interpolated map.

2.1.1. Bias

Linearity is very convenient for a study of the bias of the smoothing, since it allows us to write the average value of a smoothed function $\tilde{f}$ as

$$
\langle \tilde{f}(x) \rangle = \int_X f(\bar{x}) K(x; \bar{x}) \rho(\bar{x})\,d\bar{x}. \tag{14}
$$

In other words, the bias of the estimator is completely described by the kernel function $K(x; \bar{x})$. Note also that the only statistical property of measurement errors $\{\epsilon_i\}$ that have been used to derive Eq. (14) is $\langle \epsilon_i \rangle = 0$; again, this is due to linearity. Almost all interpolators considered in this paper are linear.

In order to show Eq. (14), we consider the average $\langle \tilde{f}(x) \rangle$ and use the linearity relation (10):

$$
\begin{aligned}
\langle \tilde{f}(x) \rangle &= \left\langle S(x; \{(x_i, f(x_i) + \epsilon_i)\}) \right\rangle \\
&= \left\langle \sum_{j=1}^{N} [f(x_j) + \epsilon_j] S(x; \{(x_i, \delta_{ij})\}) \right\rangle.
\end{aligned}
\tag{15}
$$

We consider now the average on errors, and note that the ϵ_j can be dropped from the previous equation because $\langle \epsilon_j \rangle = 0$. Let us now decompose f in the form

$$
f(x) = \int_X \delta(x - \bar{x}) f(\bar{x})\,d\bar{x}, \tag{16}
$$

where δ is Dirac's distribution. Inserting this expression into Eq. (15) we obtain

$$
\begin{aligned}
\langle \tilde{f}(x) \rangle &= \left\langle \int_X d\bar{x}\, f(\bar{x}) \sum_{j=1}^{N} \delta(x_j - \bar{x}) S(x; \{(x_i, \delta_{ij})\}) \right\rangle \\
&= \int_X f(\bar{x}) K(x; \bar{x}) \rho(\bar{x})\,d\bar{x},
\end{aligned}
\tag{17}
$$

where $K(x; \bar{x})$ is given by

$$
\begin{aligned}
K(x; \bar{x}) &= \frac{1}{\rho(\bar{x})} \left\langle \sum_{j=1}^{N} \delta(x_j - \bar{x}) S(x; \{(x_i, \delta_{ij})\}) \right\rangle \\
&= \frac{1}{\rho(\bar{x})} \left\langle S(x; \{(x_i, \delta(x_i - \bar{x}))\}) \right\rangle.
\end{aligned}
\tag{18}
$$

Note that we have defined $K(x; \bar{x})$ with a factor $1/\rho(\bar{x})$ for further convenience. Equation (17) is precisely Eq. (14).

The practical evaluation of the kernel K is very important for a statistical study of the interpolation technique. Indeed, this kernel controls the relationship between the expected interpolated map and the original, unknown field, and thus is useful, for example, to perform a comparison between the observations and some theoretical expectation. In order to evaluate K, we first observe that almost all linear interpolation procedures have a simple property: the interpolated value at x does not strongly depend on values that are far away from it. In this case, the kernel K associated with a linear interpolator can be obtained using a numerical technique that we now describe (a proof will

follow). The procedure used in order to determine $K(x; \bar{x})$ is described in the following items:

1. Choose a large subset $A \subset X$ that contains both x and $\bar{x}$.
2. Generate N, the number of objects inside A, according to the scheme chosen for the distribution of the locations (for example, if the locations are independent and uniformly distributed with density ρ, then N follows the Poisson distribution given in Eq. (7)).
3. Generate the N locations $\{x_i\}$ inside A following the scheme chosen (for example, uniformly if the density is constant and there is no correlation).
4. Assign a vanishing value $y_i = 0$ to each location x_i. Assign a value 1 to an extra location at $\bar{x}$.
5. The configuration considered is composed by the object at $\bar{x}$ and the N objects at $\{x_i\}$; hence the configuration is $\Upsilon = \{(\bar{x}, 1)\} \cup \{(x_i, 0)\}$.
6. Evaluate $S(x; \Upsilon)$ if this function is defined at x; otherwise arbitrarily assign a vanishing value to the interpolated value at x.
7. Generate several configurations by repeating points 2–6, and evaluate the average of $S(x; \Upsilon)$. This average, multiplied by $[1 - P_0(x)]^{-1}$, is an estimate of $K(x; \bar{x})$.

This procedure will be used in this paper not only to derive numerically the kernels of several interpolating procedures, but also to obtain some general properties of interpolators.

A proof of this practical method can be obtained as follows. From Eq. (18) we see that, in principle, K could be evaluated by performing an integration over the probability distribution function for $\{x_i\}$ using the whole set X. In practice, this method could never be applied in numerical studies because of the presence of δ Dirac's distribution and because of the large number of locations to be generated (actually, since X is not bounded we should generate an infinite number of objects). However, a different approach is feasible. First, we use a Monte-Carlo integration, that is, we generate a set of locations $\{x_i\}$ according to the expected probability distribution. If we assume that the smoothing is weakly dependent on values y_i whose locations x_i are far away from x, we can safely generate points inside a subset $A \subset X$ that abundantly contains x and $\bar{x}$. Then, in order to solve the problem with Dirac's delta, we approximate this distribution with a top-hat function $H(\bar{x} - x_i)$ centered on $\bar{x}$ and with area s. In other words, we take $H(\bar{x} - x_i)$ to be $1/s$ if x_i falls inside a region of area s around $\bar{x}$, and we take 0 otherwise; we will eventually let s go to zero. Since s is taken to be small, in most cases all locations $\{x_i\}$ will not be close enough to $\bar{x}$ and thus all values $\{y_i\}$ will vanish. In such situations, since the interpolator is linear, we obtain a vanishing value at x. In a few cases, however, we expect to have configurations with a single point inside s, and thus with value $1/s$; such configurations have probability $\rho(\bar{x})s$ as $s \to 0$. When we now take the average of S, cases where all values vanish do not contribute to the average because the interpolated value vanishes as well; other cases, instead, contribute with a term proportional to $1/s$ (the value corresponding to the location around $\bar{x}$). Equivalently, we can estimate the relevant average forcing a point inside the top-hat close to $\bar{x}$ and multiplying the resulting average by the probability

that this happens, i.e. $\rho(\bar{x})s$. Since the interpolator is linear, this is equivalent to forcing a point with value 1 at $\bar{x}$, taking the average, and multiplying the result by $\rho(\bar{x})$; this last factor, however, actually cancels with the term $1/\rho(\bar{x})$ used in the definition of $K(x; \bar{x})$ (cf. Eq. (18)). This proves the correctness of the procedure to obtain K described in the points above.

In our discussion we have implicitly assumed that the estimator S is always defined, i.e. that $P_0(x) = 0$. If this is not the case, we must slightly modify the method described above. In particular, we need to distinguish between four probabilities: $P_{s,1}$, the probability of having a point inside s and the smoothing procedure defined at x, $P_{\mathit{s},1}$, the probability of having no point inside s and the smoothing procedure defined at x, and the two symmetrical probabilities $P_{s,0}$ and $P_{\mathit{s},0}$ for configurations where the interpolator is not defined at x. Note that $P_{s,0} + P_{\mathit{s},0} = P_0(x)$ as defined in Sect. 1.1. We then find that the procedure described above can be still applied if $P_0(x) \neq 0$, provided that two changes are made:

– we need to discard, in the ensemble average of points, configurations for which S cannot be evaluated at x.
– we need to multiply the average of S obtained from Monte-Carlo integration by the probability that, in the ensemble average, a point falls inside s. In the general case $P_0(x) \neq 0$ considered here, this probability is given by $P_{s,1}/[s(P_{s,1} + P_{\mathit{s},1})]$.

Both problems can be solved at the same time using the following procedure. First, we note that the if we set $y = 0$ for cases where S cannot be evaluated at x, we are basically multiplying the average by a factor $P_{s,1}/(P_{s,0} + P_{s,1})$. On the other hand, we know from the previous analysis that $P_{s,0} + P_{s,1} = \rho(\bar{x})s$. Hence, we can recover the correct factor by setting $y = 0$ if S is not defined and by multiplying the final result by $[1 - P_0(x)]^{-1}$. This completes the proof for $K(x; \bar{x})$.

2.1.2. Covariance

Linearity has also important consequences for the form of the covariance. In particular, the term $\langle \tilde{f}(x) \tilde{f}(x') \rangle$ can be written as

$$\langle \tilde{f}(x) \tilde{f}(x') \rangle = \int_X d\bar{x} \rho(\bar{x}) C_1(x, x'; \bar{x})[\sigma^2 + f(\bar{x})f(\bar{x})]$$
$$+ \int_X d\bar{x} \rho(\bar{x}) \int_X d\bar{x}' \rho(\bar{x}') C_2(x, x'; \bar{x}, \bar{x}') f(\bar{x}) f(\bar{x}'). \quad (19)$$

Hence, the covariance is composed by a term proportional to σ^2, the scatter of measurements $\{y_i\}$ around the true values $\{f(x_i)\}$, and additional terms that can be identified as Poisson noise. More specifically, the two noise terms are (cf. Eq. (6))

$$T_\sigma(x, x') = \sigma^2 \int_X d\bar{x} \rho(\bar{x}) C_1(x, x'; \bar{x}), \quad (20)$$

$$T_P(x, x') = \int_X d\bar{x} \rho(\bar{x}) C_1(x, x'; \bar{x})[f(\bar{x})]^2$$
$$+ \int_X d\bar{x} \rho(\bar{x}) \int_X d\bar{x}' \rho(\bar{x}') C_2(x, x'; \bar{x}, \bar{x}') f(\bar{x}) f(\bar{x}')$$
$$- \int_X d\bar{x} \rho(\bar{x}) \int_X d\bar{x}' \rho(\bar{x}') K(x; \bar{x}) K(x'; \bar{x}') f(\bar{x}) f(\bar{x}').$$
$$(21)$$

Again, the fact that measurement errors enter only through σ^2 is a consequence of linearity.

In order to show Eq. (19), we use again the linearity of S to write (cf. Eq. (15))

$$\tilde{f}(x) = S(x; \{(x_i, f(x_i) + \epsilon_i)\})$$
$$= \sum_{j=1}^{N} [f(x_j) + \epsilon_j] S(x; \{(x_i, \delta_{ij})\}). \tag{22}$$

Using this expression in the product $\langle \tilde{f}(x)\tilde{f}(x') \rangle$, we obtain

$$\langle \tilde{f}(x)\tilde{f}(x') \rangle = \left\langle \sum_{j,j'} [f(x_j) + \epsilon_j][f(x_{j'}) + \epsilon_{j'}] \right.$$
$$\times S(x; \{(x_i, \delta_{ij})\}) S(x'; \{(x_{i'}, \delta_{i'j'})\}) \right\rangle$$
$$= \left\langle \sum_{j} [f(x_j)f(x_j) + \sigma^2] S(x; \{(x_i, \delta_{ij})\}) \right.$$
$$\times S(x'; \{(x_{i'}, \delta_{i'j})\}) \right\rangle$$
$$+ \left\langle \sum_{j \neq j'} f(x_j)f(x_{j'}) S(x; \{(x_i, \delta_{ij})\}) \right.$$
$$\times S(x'; \{(x_{i'}, \delta_{i'j'})\}) \right\rangle. \tag{23}$$

Note that we have explicitly separated the cases $j = j'$ and $j \neq j'$ in order to take advantage of Eq. (4). We now use again the decomposition (16) for $f(x)$, $[f(x)]^2$, and σ^2 (the last taken to be a constant function of x), thus obtaining

$$\langle \tilde{f}(x)\tilde{f}(x') \rangle = \int d\bar{x} \rho(\bar{x})[f(\bar{x})f(\bar{x}) + \sigma^2]C_1(x, x'; \bar{x})$$
$$+ \int d\bar{x} \rho(\bar{x}) \int d\bar{x}' \rho(\bar{x}')f(\bar{x})f(\bar{x}')C_2(x, x'; \bar{x}, \bar{x}'), \tag{24}$$

with

$$C_1(x, x'; \bar{x}) = \frac{1}{\rho(\bar{x})} \left\langle \sum_{j} \delta(\bar{x} - x_j) S(x; \{(x_i, \delta_{ij})\}) \right.$$
$$\times S(x'; \{(x_{i'}, \delta_{i'j})\}) \right\rangle, \tag{25}$$

$$C_2(x, x'; \bar{x}, \bar{x}') = \frac{1}{\rho(\bar{x})\rho(\bar{x}')} \left\langle \sum_{j \neq j'} \delta(\bar{x} - x_j)\delta(\bar{x}' - x_{j'}) \right.$$
$$\times S(x; \{(x_i, \delta_{ij})\}) S(x'; \{(x_{i'}, \delta_{i'j'})\}) \right\rangle. \tag{26}$$

Again, the factors $\rho(\bar{x})$ and $\rho(\bar{x}')$ have been introduced here to simplify some of the following equations. This proves Eq. (19). Note that, although Eq. (26) is apparently composed of two independent factors, $S(x; \{(x_i, \delta_{ij})\})$ and $S(x'; \{(x_{i'}, \delta_{i'j'})\})$, in reality both terms are functions of *all* locations $\{x_i\}$; hence, since the random variables that enters the expression for C_2 are precisely the locations $\{x_i\}$, the two terms are correlated and Eq. (26) cannot further simplified. The same, clearly, applies to Eq. (25).

Similarly to the average, the two kernels C_1 and C_2 can be numerically evaluated using a simple procedure that we now describe:

1. Choose a large subset $A \subset X$ that contains x, $\bar{x}$, x', and $\bar{x}'$.

2. Generate N, the number of objects inside A, according to the scheme chosen for the distribution of the locations (for example, if the locations are independent and uniformly distributed with density ρ, then N follows the Poisson distribution given in Eq. (7)).

3. Generate the N locations $\{x_i\}$ inside A following the scheme chosen (for example, uniformly if the density is constant and there is no correlation).

4. Assign a vanishing value $y_i = 0$ to each location x_i. For C_1, assign a value 1 to an extra location at $\bar{x}$, similarly to what we do for K. For C_2, assign a value 1 to two extra locations at $\bar{x}$ and $\bar{x}'$.

5. Hence, the two configurations considered are $\Upsilon_1 = \Upsilon = \{(\bar{x}, 1)\} \cup \{(x_i, 0)\}$ and $\Upsilon_2 = \{(\bar{x}, 1), (\bar{x}', 1)\} \cup \{(x_i, 0)\}$.

6. For C_1 evaluate $S(x; \Upsilon_1)S(x'; \Upsilon_1)$, if the function S is defined at both points x and x'; otherwise arbitrarily assign a vanishing value to the product above. Analogously, for C_2 evaluate $S(x; \Upsilon_2)S(x'; \Upsilon_2)$, if this product is defined, or use zero otherwise.

7. Repeat points 2–6, and evaluate the average $\langle S(x; \Upsilon_1)S(x'; \Upsilon_1) \rangle$. This average, multiplied by $[1 - P_0(x, x')]^{-1}$, is an estimate of $C_1(x, x'; \bar{x})$. An estimate of $C_2(x, x'; \bar{x}, \bar{x}')$ is instead given by the average $\langle S(x; \Upsilon_2)S(x'; \Upsilon_2) \rangle$ multiplied by $[1 - P_0(x, x')]^{-1}$.

Again, the numerical techniques described in the items above will be also used to derive some general properties for the covariance of interpolators in the following sections.

This practical method to evaluate C_1 and C_2 can be derived using a technique similar to the one described for K. In particular, Eq. (25) suggests that we can determine the kernel C_1 as follows. We approximate the delta distribution in $\bar{x}$ with a top-hat function $H(\bar{x} - x_j)$ of area s. Then, we randomly generate the locations inside a large subset $A \subset X$ that contains x, x', and $\bar{x}$. As for the kernel K, most configurations will have only points outside the top-hat, thus resulting in a vanishing sum. A fraction $s\rho(\bar{x})$ of configurations, however, will have a single point falling inside the top-hat function, i.e. with coordinate very close to $\bar{x}$. Since the configurations without points inside s do not contribute to the average, we can just ignore them and consider only configurations with a point at $\bar{x}$; we then multiply the average by the probability of having such configurations, i.e. $\rho(\bar{x})s$ in the limit $s \to 0$. Since the function H has a value $1/s$ at $\bar{x}$, this is equivalent to assigning a value 1 to the point at $\bar{x}$, and multiply the final result by $\rho(\bar{x})$; this term, however, disappears because of the presence of a factor $1/\rho(\bar{x})$ in the definition of C_1.

Again, in this procedure, we have ignored cases where the smoothing technique cannot be applied. In order to evaluate the expression in the right hand side of Eq. (25), we need the values of S at x and x'. Hence, we should discard cases where S cannot be evaluated at x or x'. Alternatively, we can proceed as done for K, and just assign, for the configuration Υ, vanishing value to the product $S(x; \Upsilon)S(x'; \Upsilon)$ when one of the two factors cannot be evaluated. Finally, we multiply the total result by a factor $[1 - P_0(x, x')]^{-1}$, where this extra term is used to correctly normalize the average of Eq. (25).

The evaluation for C_2 is similar. In this case, however, we have two different delta distributions at $\bar{x}$ and $\bar{x}'$. Hence, we approximate them with two top-hat functions $H(\bar{x} - x_j)$ and $H(\bar{x}' - x'_j)$, both of area s. Since we are going to take the limit $s \to 0$, the two top-hat functions do not intersect, and thus the probability of having a point inside both top-hats is the product of the individual probabilities, i.e. $s^2 \rho(\bar{x})\rho(\bar{x}')$. Then, in the limit $s \to 0$, we can force a point at $\bar{x}$ and one at $\bar{x}'$, both with value 1; other points are randomly distributed with density ρ. Then, we evaluate the average of $S(x; \Upsilon)S(x; \Upsilon)$ and multiply the final result by $\rho(\bar{x})\rho(\bar{x}')$; this factor, however, disappears from C_2 (cf. prefactor in Eq. (26)) and is shifted instead in the integration (24). Again, if P_0 is not vanishing, we need to correct for cases where S cannot be evaluated at x or x'; the correcting factor is still $1/[1 - P_0(x, x')]$.

2.2. Normalization

Almost every smoothing procedure used in astronomy has a simple normalization property: if all values are the same, i.e. $y_i = c$, then the interpolating function S always returns c at every point x:

$$S(x; \{(x_i, c)\}) = c. \tag{27}$$

Normalized smoothing procedures are sometimes called "unbiased". This property is satisfied, for example, by the interpolator (2). For a linear smoothing, this property immediately implies

$$\int_X K(x; \bar{x})\rho(\bar{x})\, d\bar{x} = 1 \qquad \forall\, x \in X. \tag{28}$$

This can be easily verified by using a constant function $f(x) = 1$ and by noting that in this case $\langle \tilde{f}(x) \rangle = 1$; Eq. (14) then gives the normalization of K. Similarly, for the covariance kernels we have

$$\int_X d\bar{x}\, \rho(\bar{x})C_1(x, x'; \bar{x})$$
$$+ \int_X d\bar{x}\, \rho(\bar{x}) \int_X d\bar{x}'\, \rho(\bar{x}')C_2(x, x'; \bar{x}, \bar{x}') = 1. \tag{29}$$

Note that the Poisson noise T_P vanishes if $f(x)$ is "flat."

Given any linear smoothing procedure S written in the form (12), we can obtain a related normalized interpolator S':

$$S'(x; \{(x_i, y_i)\}) = \left[\sum_{j=1}^{N} w_j(x; \{x_i\})y_j \right]$$
$$\times \left[\sum_{j=1}^{N} w_j(x; \{x_i\}) \right]^{-1}. \tag{30}$$

Indeed, in many cases interpolating techniques are directly written as in Eq. (30). An example is given by our toy-interpolator (2), which is of the form (30) with

$$w_j(x; \{x_i\}) = 1/|x - x_j|^\alpha. \tag{31}$$

2.3. Spatial symmetries

Basically all interpolation procedures of some interest satisfy some spatial symmetries. In this section we will consider three common spatial invariance properties, namely translation, rotation, and scaling invariance. These properties can have important consequences on the forms of the three kernels K, C_1, and C_2, provided that similar symmetries applies for the spatial distribution of location measurements.

2.3.1. Invariance upon translation

Many smoothing techniques are invariant upon translation, i.e. there is no "preferred" point on X for the smoothing:

$$S(x + d; \{(x_i + d, y_i)\}) = S(x; \{(x_i, y_i)\}). \tag{32}$$

For example, the simple smoothing (2) is invariant upon translation because it depends only on the distances $|x - x_i|$ between the interpolated point x and the locations $\{x_i\}$. A linear smoothing invariant upon translation has necessarily associated weights w_i (cf. Eq. (12)) of the form $w_i = w_i(\{x_j - x\})$. Moreover, if the distribution of locations is also invariant upon translation (this implies, among other things, that the density ρ is uniform), then the kernel K is also invariant upon translation, i.e. $K(x; \bar{x}) = K(x - \bar{x})$, and thus Eq. (14) becomes a simple convolution.

If the interpolation procedure is also normalized, than an interesting property holds:

$$\int_X \langle \tilde{f}(x) \rangle\, dx = \int_X f(x)\, dx. \tag{33}$$

This can be shown by noting that for a smoothing procedure invariant upon translation, $\langle \tilde{f} \rangle$ is the convolution of f with the kernel K. Moreover, for a normalized smoothing the kernel K is normalized to unity (cf. Eq. (28)), and thus Eq. (33) holds. Note that this result basically states the conservation of the "signal": the total measured signal is equal to the true, original signal.

The invariance upon translation has also consequences on the two covariance kernels. More specifically, they must be of the form $C_1(x, x'; \bar{x}) = C_1(x - \bar{x}, x' - \bar{x})$ and $C_2(x, x'; \bar{x}, \bar{x}') = C_2(x - \bar{x}, x' - \bar{x}; \bar{x}' - \bar{x})$. Note also that $P_0(x)$ is independent of x; hence, we will write just P_0; similarly, $P_0(x, x')$ can depend only on $(x - x')$.

2.3.2. Invariance upon rotation

Smoothing procedures in spaces X of dimension larger than one are also often invariant upon rotation and mirror symmetry, i.e. they are isotropic. Formally

$$S(x; \{(x + R(x_i - x), y_i)\}) = S(x; \{(x_i, y_i)\}). \tag{34}$$

In this equation, R is any orthogonal matrix; note that the rotation is actually performed around the point x. The smoothing (2) is invariant upon rotation (to show this we note, again, that this interpolator only depends on the distances $|x - x_i|$).

If this symmetry holds also for the generation of locations (this implies, again, that ρ is uniform) and if the interpolator is linear, than the kernel K is of the form $K(x; \bar{x}) = K'(x; |x - \bar{x}|)$,

i.e. the smoothing kernel can depend on x and on the distance $|x - \bar{x}|$ between x and $\bar{x}$ only.

The kernel associated with an interpolator invariant upon rotation has an interesting property. Let us evaluate the integral

$$\int_X (x - \bar{x}) K(x; \bar{x}) \, d\bar{x} = \int_X (x - \bar{x}) K(x; |x - \bar{x}|) \, d\bar{x} = 0 , \quad (35)$$

where the last equality holds because the integrand is an odd function of $(x - \bar{x})$. We can recast this result in a more interesting form

$$\left[\int_X \bar{x} K(x; \bar{x}) \, d\bar{x} \right] \left[\int_X K(x; \bar{x}) \, d\bar{x} \right]^{-1} = x . \quad (36)$$

This quantity in the left hand side of this equation is a measure of the "center of mass" for the kernel K; hence Eq. (36) basically assures that there is no systematic "offset" on the final map.

If both spatial invariance properties considered so far hold for a linear smoothing procedure, then the kernel K associated with the smoothing is of the form $K(x; \bar{x}) = K''(|x - \bar{x}|)$, so that only distances are involved. This property also holds for the covariance kernels, which are only dependent on the various distances between x, x', $\bar{x}$, and (for C_2 alone) $\bar{x}'$. As a result, the noise term T_σ is a function of $|x - x'|$ alone.

2.3.3. Invariance upon scaling

A third spatial invariance occurs for some smoothing techniques, which are intrinsically scale-free: the result of the smoothing depends only on the *ratios* of distances and not on their absolute values. Formally in terms of the smoothing function, we have

$$S(x; \{(x + k(x_i - x), y_i)\}) = S(x; \{(x_i, y_i)\}), \quad (37)$$

where k is an arbitrary positive real number. It is not difficult to verify that the interpolator (2) is invariant upon scaling. Indeed, for any positive real k, we have

$$S(x; \{(x + k(x_i - x), y_i)\}) = \frac{\sum_{j=1}^N y_j / (k^\alpha |x - x_j|^\alpha)}{\sum_{j=1}^N 1 / (k^\alpha |x - x_j|^\alpha)}$$

$$= \frac{\sum_{j=1}^N y_j / |x - x_j|^\alpha}{\sum_{j=1}^N 1 / |x - x_j|^\alpha} = S(x; \{(x_i, y_i)\}). \quad (38)$$

If measurement locations are uncorrelated and uniformly distributed (i.e., for a homogeneous Poisson process), then the scale-free property allows us to derive a number of consequences that greatly simplify the analysis of the estimator. For a scale-free linear smoothing, the typical scale-length of the smoothing kernel K is set only by the density of objects ρ (no other scales are available; note instead that in presence of clustering we would immediately have another scale, the correlation length). Formally, this can be seen by noting that two different location configurations, $\{x_i\}$ and $\{x + k(x_i - x)\}$ have the same probability if the density is changed from ρ to $k^{-n}\rho$,

where n is the space dimension. As a result, all kernels are simply influenced by a change of ρ:

$$\rho \mapsto k^{-n}\rho, \quad (39)$$

$$K(\bar{x} + d; \bar{x}) \mapsto K(\bar{x} + kd; \bar{x}), \quad (40)$$

$$C_1(\bar{x} + d, \bar{x} + d'; \bar{x}) \mapsto C_1(\bar{x} + kd, \bar{x} + kd'; \bar{x}), \quad (41)$$

$$C_2(\bar{x} + d, \bar{x} + d'; \bar{x}, \bar{x} + \bar{d}) \mapsto$$
$$C_2(\bar{x} + kd, \bar{x} + kd'; \bar{x}, \bar{x} + k\bar{d}). \quad (42)$$

Hence, we can study the kernels associated to a smoothing technique invariant upon scaling for a given density (say $\rho = 1$) and then generalize the results obtained using Eqs. (39)–(42). We also note that for a normalized smoothing, as $\rho \to \infty$ the product $\rho(\bar{x})K$ will be proportional to Dirac's delta distribution.

2.4. Exact interpolators

An interpolating procedure is said to be exact if it honors the data points upon which the interpolation is based:

$$S(x_j; \{(x_i, y_i)\}) = y_j. \quad (43)$$

Honoring data points is seen as an important feature in many applications; on the other hand, if it is known that the data points might be affected by significant errors, exact interpolators might not be the right choice.

Regarding the exactness, the interpolator (2) needs a few words of explanation. Indeed, when the interpolated position x coincides with one of the locations $\{x_i\}$, the expression (2) becomes singular, and thus the interpolator is apparently undefined. However, the limit $x \to x_i$ of S is defined and equal to y_i (we assume here that all locations are different; the case of two or more coincident locations has no statistical relevance). Hence, we can safely extend the definition of the smoothing technique (2) to all points $x \in X$, with the convention that limits must be taken if the expression is undefined. This way, we can state that our toy-interpolator is exact.

Recalling the numerical method used to evaluate K for a linear interpolator, we can deduce that an exact estimator has the property $K(\bar{x}; \bar{x}) = 1/[1 - P_0(\bar{x})]$. In fact, if we use the numerical method, we will always measure $S(\bar{x}; \{(x_i, y_i)\}) = 1$, since we have assigned a value 1 to the location $\bar{x}$, and thus the average of all those measurements is also 1 (here we are neglecting cases where the smoothing is not defined). Similarly, we find $C_1(\bar{x}, \bar{x}; \bar{x}) = 1/[1 - P_0(\bar{x}, \bar{x})]$ and $C_2(\bar{x}, \bar{x}'; \bar{x}, \bar{x}') = C_2(\bar{x}', \bar{x}; \bar{x}, \bar{x}') = 1/[1 - P_0(\bar{x}, \bar{x}')]$.

2.5. Bounded interpolators

An interpolating procedure is said to be bounded if interpolated values are always between the smallest and the largest measured values:

$$\min y_i \le S(x; \{(x_i, y_i)\}) \le \max y_i. \quad (44)$$

This is a rather natural property satisfied by many smoothing techniques (e.g., the interpolator (2)). Note that a bounded interpolator is always normalized, since Eq. (44) implies Eq. (27).

If a linear interpolator is bounded, then we can put superior and inferior limits on $K(x; \bar{x})$. In fact, recalling again the numerical technique used to evaluate K, we obtain

$$0 \leq K(x; \bar{x}) \leq [1 - P_0(x)]^{-1}. \tag{45}$$

Hence, the kernel K is non-negative and $[1 - P_0(x)]^{-1}$ is a superior limit for it. Similarly, for the other kernels we obtain

$$0 \leq C_1(x, x'; \bar{x}) \leq [1 - P_0(x, x')]^{-1}, \tag{46}$$

$$0 \leq C_2(x, x'; \bar{x}, \bar{x}') \leq [1 - P_0(x, x')]^{-1}. \tag{47}$$

We can go further in this analysis by defining the effective number of objects used by the smoothing procedure as

$$\mathcal{N}_{\text{eff}}(x) = \frac{1}{1 - P_0(x)} \left[\int_X K(x; \bar{x}) \rho(\bar{x}) \, d\bar{x} \right]^2$$
$$\times \left[\int_X [K(x; \bar{x})]^2 \rho(\bar{x}) \, d\bar{x} \right]^{-1}. \tag{48}$$

Note that $\mathcal{N}_{\text{eff}}$ can be defined for any linear smoothing procedure; for a normalized (and in particular for a bounded) interpolator, the first term in the right hand side of Eq. (48) is unity. For an interpolator invariant upon translation, the quantity $\mathcal{N}_{\text{eff}}(x)$ does not depend on x, and thus we will write just $\mathcal{N}_{\text{eff}}$. Using the definition (48) we can recast the upper limit for $K(x; \bar{x})$, valid for a bounded interpolator, as a lower limit for the effective number of objects $\mathcal{N}_{\text{eff}}(x)$:

$$\mathcal{N}_{\text{eff}}(x) \geq \frac{1}{1 - P_0(x)} \left[\int_X \frac{1}{1 - P_0(x)} K(x; \bar{x}) \rho(\bar{x}) \, d\bar{x} \right]^{-1}$$
$$\geq 1. \tag{49}$$

Equation (49) put a lower limit to the effective number for the kernel K, i.e. to the expected *resolution* of the smoothing. Finally, we observe that interpolators for which the scaling invariance holds have effective number independent of ρ (remember that ρ is supposed to be constant for scaling invariant interpolators).

The covariance of a bounded interpolating procedure satisfies some interesting properties. Using $C_1 \geq 0$ (see (46)) in Eq. (29), in fact, we obtain

$$\int_X d\bar{x} \rho(\bar{x}) C_1(x, x'; \bar{x}) \leq 1, \tag{50}$$

so that $T_\sigma(x, x') \leq \sigma^2$. Hence, we have obtained an upper limit for the contribution to the noise from measurement errors. A comparison of this result with the inequality $\mathcal{N}_{\text{eff}} \geq 1$ suggests that our definition of effective number of objects is sensible.

We can better appreciate the relationship between the covariance kernel C_1 and $\mathcal{N}_{\text{eff}}$ with the following argument, valid for any linear interpolation procedure. Let us consider $C_1(x, x; \bar{x})$ and briefly recall how this quantity can be measured using the numerical approach discussed in Sect. 2.1.2. We generate locations, keeping one object with weight 1 in $\bar{x}$, and take the average of the quantity $S(x; \Upsilon_1) S(x; \Upsilon_1)$; finally we multiply the average by $1/[1 - P_0(x)]$, because in this particular case $P_0(x, x) = P_0(x)$. Something quite similar is done for $K(x; \bar{x})$ (see Sect. 2.1.1): the configurations used are the same,

in the sense that $\Upsilon = \Upsilon_1$, but in this case we take the average of $S(x; \Upsilon)$. From this discussion, we see that $K(x; \bar{x})$ is the simple average of a set of some non-negative numbers, while $C_1(x, x; \bar{x})$ is, apart from a known numerical factor, the average of the squares of the same quantities. Hence, the following inequality holds:

$$C_1(x, x; \bar{x}) \geq [1 - P_0(x)][K(x; \bar{x})]^2. \tag{51}$$

For a normalized interpolator, this equation implies $T_\sigma(x, x) \geq \sigma^2 / \mathcal{N}_{\text{eff}}$. Hence, the effective number of objects puts a lower limit on the measurement error.

2.6. Local interpolators

An interpolator is called local if the interpolated value at x depends only on the values y_i for locations x_i "close" to x; an interpolator which is not local is said to be global. Global interpolators usually produce smoother maps but often require more computing time. The interpolator (2) is global.

We further distinguish between strongly and weakly local interpolators depending on the meaning that is given to the word "close" in the previous definition. An interpolator is strongly local if it uses only values y_i whose locations x_i fall within a fixed range from x; a weakly local interpolator, instead, is only guaranteed to use a *finite* number of points close to x. We stress that this definition is not standard but is needed for a rigorous characterization of local smoothing techniques. Note that a strongly local interpolator cannot be scale invariant, because it depends only on points within a *fixed range* and hence, there is a natural scale for the interpolator, its range. A strongly local interpolator has always $P_0(x) > 0$, because the interpolator cannot be defined if there is a large "void" without points around x.

Using again the numerical technique to obtain K, we immediately find that the kernel K associated with a strongly local interpolator has compact support, i.e. $K(x; \bar{x})$ vanishes if $|x - \bar{x}|$ is large; analogously, $C_1(x, x'; \bar{x})$ vanishes if either $|x - \bar{x}|$ or $|x' - \bar{x}|$ are large, and $C_2(x, x'; \bar{x}, \bar{x}')$ vanishes if x or x' are far away from $\bar{x}$ or $\bar{x}'$. Note also that for strong local estimators, if $|\bar{x} - \bar{x}'|$ is large, $C_2(x, x'; \bar{x}, \bar{x}')$ converges to

$$C_2(x, x'; \bar{x}, \bar{x}') = \frac{[1 - P_0(x)][1 - P_0(x')]}{[1 - P_0(x, x')]}$$
$$\times [K(x; \bar{x}) K(x'; \bar{x}') + K(x'; \bar{x}) K(x; \bar{x}')]. \tag{52}$$

Not much, instead, can be said regarding weakly local interpolators; indeed, weakly local interpolators have been defined here mainly to have a complete classification of locality.

Several strongly local interpolators can be defined with just a single point inside their range. In this case, we can show that in the limit $\rho \to 0$, the kernel K associated to a interpolator approaches a top-hat function, provided the density ρ is uniform. For if ρ is sufficiently small everywhere, we expect a negligible probability, when using the numerical technique to obtain $K(x; \bar{x})$, of having points other than $\bar{x}$ inside the range of the function. In this case, all terms in the sums of Eq. (30) vanish except for $\bar{x}$. As a result, we always measure 1 for points inside the range of the interpolator, and 0 outside. Moreover, if π is

set around the point x where the interpolator is defined, we find (cf. Eq. (68))

$$P_0(x) = \exp\left(-\int_\pi \rho(x')\,dx'\right) \simeq 1 - \int_\pi \rho(x')\,dx'. \qquad (53)$$

Notice that the dependence on x of $K(x;\bar{x})$ is also expressed by the set π, which is a function of the location x. In summary, we obtain

$$K(x;\bar{x}) \simeq \begin{cases} 1/\int_\pi \rho(x')\,dx' & \text{if } \bar{x} \in \pi, \\ 0 & \text{otherwise.} \end{cases} \qquad (54)$$

As expected, this kernel is manifestly normalized according to Eq. (27). In case of a uniform distribution ρ, this expression simplifies to

$$K(x;\bar{x}) \simeq \begin{cases} 1/\rho\mu(\pi) & \text{if } \bar{x} \in \pi, \\ 0 & \text{otherwise,} \end{cases} \qquad (55)$$

where $\mu(\pi)$ is the area of the set π. Hence, in the limit $\rho \to 0$ we reach the upper bound for K stated in Eq. (45). Regarding the covariance kernel, calling $\mu(\pi \cap \pi')$ the area of the intersection of the ranges for x and x', we obtain (here, for simplicity, we assume a uniform location density ρ)

$$C_1(x,x';\bar{x}) \simeq \begin{cases} 1/\rho\mu(\pi \cap \pi) & \text{if } \bar{x} \in \pi \cap \pi\,, \\ 0 & \text{otherwise,} \end{cases} \qquad (56)$$

$$C_2(x,x';\bar{x},\bar{x}') \simeq 0. \qquad (57)$$

Note that C_2 vanishes in the limit $\rho \to 0$ (this is due to the extra factor ρ, with respect to C_1, present in the evaluation of C_2). Note also that in this case we have reached the limit $T_\sigma = \sigma^2$ discussed after Eq. (50). Finally, we can verify without difficulties that the normalization of Eq. (29) is satisfied.

2.7. Other properties

Recalling again the numerical technique to obtain K described in Sect. 2.1.1, we can consider a number of other classifications all focused on properties of S when all but one point are vanishing.

For example, suppose that the smoothed map S for the configuration $\Upsilon = \{(\bar{x}, 1)\} \cup \{(x_i, 0)\}$ is smaller than unity at each point. Such a smoothing procedure could not be bounded, but we can still obtain the upper limit $K(x;\bar{x}) < 1/(1 - P_0(x))$ for $K(x;\bar{x})$ (as in case of bounded interpolators) and Eq. (49) holds.

Another interesting property to consider is monotonicity. Suppose that, for the same configuration Υ considered above, the function S is monotonically decreasing as $|x - \bar{x}|$ increases: then $K(x;\bar{x})$ will also decrease monotonically with $|x - \bar{x}|$. If the interpolator is exact, then, this property implies $K(x;\bar{x}) \leq 1/[1 - P_0(x)]$, because $K(x,x) = 1/[1 - P_0(x)]$, and again we recover Eq. (49).

3. Kernels of some smoothing procedures

In this section we will consider several smoothing procedures and derive, analytically or numerically, the relative kernels K; for simplicity, we will not consider the kernels C_1 and C_2;
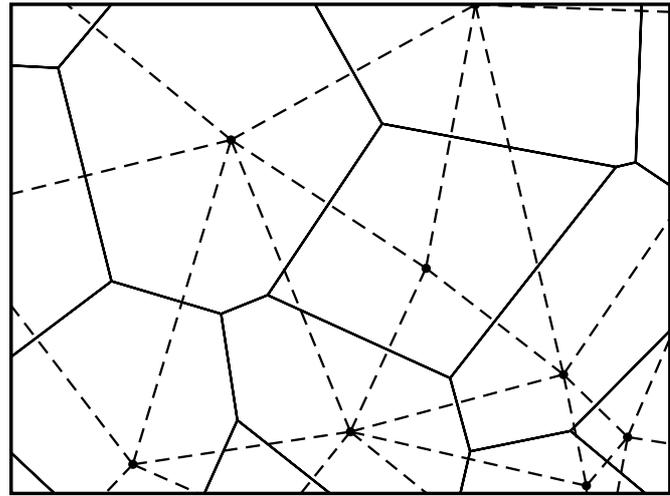
Fig. 1. The relationship between Voronoi cells (solid lines) and Delaunay triangles (dashed lines). The figure has been made by calculating the Delaunay triangulation and Voronoi tessellation for a large set of points on the plane, and by showing a small region. Note that two points are connected in the Delaunay triangulation if and only if their Voronoi cells have a side in common.

moreover, we will assume a uniform spatial distribution of locations characterized by a density ρ and no correlation. A standard approach will be followed for each interpolator. We will first briefly introduce and define the interpolator; then we will classify it according to the nomenclature introduced in Sect. 2; finally we will evaluate the kernel K.

The interpolation techniques discussed below are generally well known and are described in any book on spatial interpolation (e.g. Watson 1992). As a reference (extremely useful also for a deep discussion on statistics of spatial data) we refer to Cressie (1993), and in particular to Sect. 5.9 of this book (see also Adler 1991; Preston 1976).
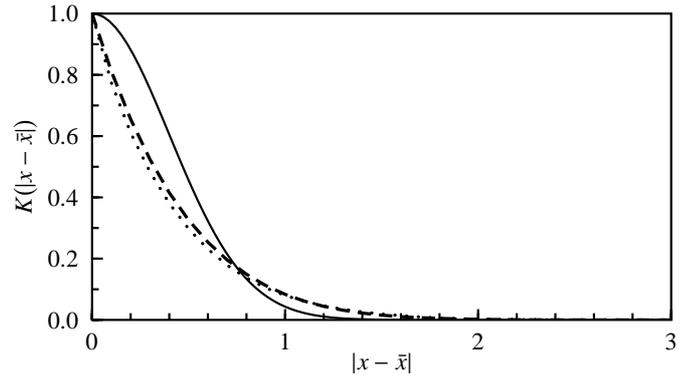
3.1. Nearest neighbor

This interpolator, called also "proximal", is probably the simplest one can imagine: the value in a point is assumed to be equal to the value of the nearest known point. Hence, the output data for such interpolator consists in Voronoi cells (see Okabe et al. 1992) centered on the points $\{x_i\}$ with abrupt changes at the boundaries (see Fig. 1).

This interpolator is manifestly linear, normalized, invariant upon translation, rotation, and scale, exact, bounded, and weakly local. As a result, we expect a kernel K of the form $K(x;\bar{x}) = K(|x - \bar{x}|)$ which scales with the density according to Eqs. (39) and (40). Moreover, we expect $P_0(x) = P_0(x, x') = 0$, $\mathcal{N}_{\text{eff}}$ independent of ρ, $K(0) = 1$, and $K(r) \leq 1$. Finally, $K(r)$ will be not increasing.

For this estimator, we can actually carry out calculations analytically using the approach discussed in Sect. 2.1.1. Let us consider a point x at distance r from $\bar{x}$. The probability that the closest location for x be $\bar{x}$ is given by $p(r) = \exp(-\rho k_n r^n)$, where n is the dimension of X and k_n is the *measure* (area or volume) of the unit sphere (e.g., $k_1 = 2$, $k_2 = \pi$, and $k_3 = 4\pi/3$).
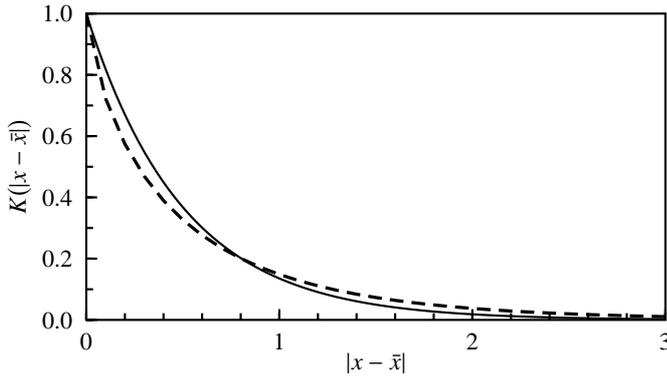
Fig. 2. The kernels K associated with the one-dimensional nearest neighbor (solid line) and Delaunay interpolators (dashed line) for $\rho = 1$. The kernel K for the natural neighbor smoothing is identical to the one for the Delaunay interpolator. Since the interpolators considered are invariant upon translation and scaling, the kernels associated to different densities can be evaluated using the transformations (39) and (40). Note also that for both plots we find $K(0) = 1$ and $K(x; \bar{x}) \leq 1$ because the smoothing techniques are exact and bounded.

In fact, $\bar{x}$ is the closest location to x if the ball of radius r centered on x does not contain any other location. The number of locations inside this ball follows a Poisson distribution with average $\rho k_n r^n$, and thus this number vanishes with the probability $p(r)$ given above. Hence, the point at x will have a value 1 with probability $p(r)$, and a vanishing value with probability $1 - p(r)$. In summary we obtain

$$K(|x - \bar{x}|) = \exp(-\rho k_n |x - \bar{x}|^n).$$ (58)

Using this expression for K we can check all the properties stated in Sect. 2 (cf. also Figs. 2 and 3). Finally, we have $\mathcal{N}_{\text{eff}} = 2$.

3.2. Delaunay triangulation

The nearest neighbor technique is very simple but unfortunately produces abrupt changes at the boundaries of the Voronoi cells, i.e. this interpolator is not smooth. This problem can be solved by using $n + 1$ near points (where n is the space dimension) and by linearly interpolating their values. For example, if $X = \mathbb{R}$, so that $n = 1$, we take the point at the left and the point at the right of x and then use linear interpolation. If $n > 1$, however, we have the problem of choosing the most convenient points to perform the linear interpolation (mathematically speaking, $\mathbb{R}^n$ is not totally ordered for $n > 1$). A standard approach is to use Delaunay triangulation (see, e.g. O' Rouche 1994), that we describe here for $n = 2$. A Delaunay triangulation for the locations $\{x_i\}$ is the unique triangulation of the plane with the property that no point in the set $\{x_i\}$ falls in the interior of the circumcircle (circle that passes through all three vertices) of any triangle in the triangulation. This definition is easily generalized for $n > 2$. The Delaunay triangulation for the points $\{x_i\}$ is also closely related to the Voronoi cells for the same points, one being the dual of the other. For example, the Delaunay triangulation can be obtained by connecting all points whose Voronoi cells have a side in common (see Fig. 1).

Fig. 3. The kernels K associated with the two-dimensional nearest neighbor (solid line), Delaunay triangulation (dashed line), and natural neighbor (dotted line) interpolators. For all plots we have used a density $\rho = 1$. Since all smoothing techniques are invariant upon translation, rotation, and scaling, the kernels associated to different densities can be evaluated using the transformations (39) and (40).

The Delaunay triangulation interpolator is thus obtained in the following way. The Delaunay triangulation is calculated for the points $\{x_i\}$; the interpolated value at x is then obtained using linear interpolation of the values for the three points that are the vertices of the triangle to which x belongs. Delaunay interpolation, or variants of it, has found already many applications in the astrophysical context (see, e.g. van de Weygaert 1994; Bernardeau & van de Weygaert 1996; Schaap & van de Weygaert 2000).

The Delaunay triangulation interpolator is linear, normalized, invariant upon translations, rotations, and scaling, exact, bounded, and weakly local. We expect for K the same properties as for the nearest neighbor.

Analytical calculations for the Delaunay triangulation linear smoothing are trivial only for dimension $n = 1$. In this case, in fact, we can just use the point at the left and the point at the right of x and proceed with linear interpolation. Calling $d_- \geq 0$ and $d_+ \geq 0$ the distances between the closest locations at the left and at the right of x and x itself, we have

$$\tilde{f}(x) = \frac{d_- f(x + d_+) + d_+ f(x - d_-)}{d_+ + d_-}.$$ (59)
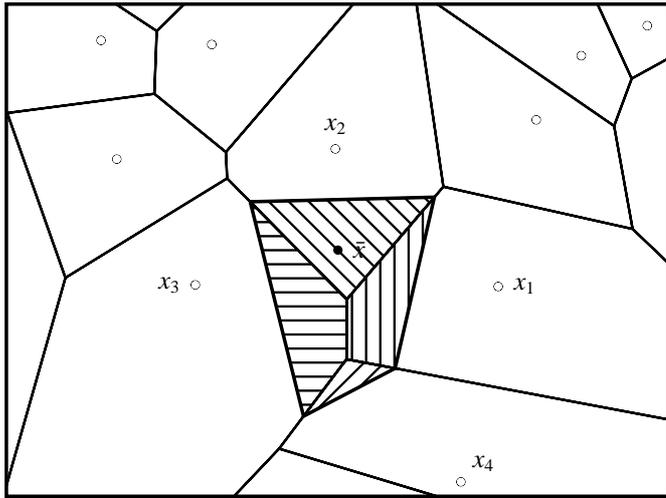
The probability distribution for both d_+ and d_- is

$$p(d_\pm) = \rho e^{-\rho d_\pm},$$ (60)

and thus we obtain

$$
\begin{aligned}
\langle \tilde{f}(x) \rangle &= \rho^2 \int_0^\infty e^{-\rho d_+} \, \mathrm{d}d_+ \\
&\quad \times \int_0^\infty e^{-\rho d_-} \frac{f(x + d_+)d_- + f(x - d_-)d_+}{d_+ + d_-} \, \mathrm{d}d_- \\
&= \rho^2 \int_{-\infty}^\infty f(x + x_1) e^{-\rho|x_1|} \, \mathrm{d}x_1 \int_0^\infty \frac{x_2 e^{-\rho x_2}}{|x_1| + x_2} \, \mathrm{d}x_2 \\
&= \rho \int_{-\infty}^\infty f(\bar{x}) K(x - \bar{x}) \, \mathrm{d}\bar{x},
\end{aligned}
$$ (61)

where the kernel K is given by

$$K(x) = e^{-\rho|x|} - \rho|x|\Gamma(0, \rho|x|).$$ (62)

Fig. 4. The evaluation of coefficients for the natural neighbor interpolation. A first Voronoi tessellation for the original set of locations (empty circles) is calculated. Then, the point x (filled circle) for which we want to obtain the interpolated value is added to the original locations, and a second Voronoi tessellation is evaluated. The new tessellation coincides with the old one, with the only difference that the new point x is now "stealing" the area for its Voronoi cell from neighboring cells (the ones corresponding to the points x_1, x_2, x_3, and x_4 in the case considered in this figure). The various "stolen areas" (marked in this figure with parallel lines of different orientation) are then used to evaluate the coefficients to be inserted in Eq. (30).

Here Γ is the incomplete gamma function, defined as

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} \, dt. \tag{63}$$

The same result is obtained using the numerical technique of Sect. 2.1.1. It can be verified that, as expected, $K(0) = 1$ and $K(x) \leq 1$; moreover, K depends only on the product $\rho|x|$, so that the scaling invariance (39–40) holds. For this kernel we find $\mathcal{N}_{\text{eff}} \simeq 2.44417$. A plot of $K(x)$ for $n = 1$ is shown in Fig. 2.

For $n > 1$ calculations cannot be easily carried out analytically. However, since the smoothing procedure is expected to be scale invariant, we can numerically evaluate K for, say, $\rho = 1$, and then apply Eqs. (39) and (40) to convert this result to different densities. Note also that, because of the translation and rotation invariance properties, K is expected to be of the form $K(|x - \bar{x}|)$, i.e. it is a function of a single real argument. Figure 3 shows the numerical results obtained for dimension $n = 2$; again, we can check that all expected properties for K are indeed satisfied. For the two-dimensional Delaunay interpolator we find $\mathcal{N}_{\text{eff}} \simeq 3.31$.

3.3. Natural neighbor

The natural neighbor interpolator (Sibson 1981), also called area stealing interpolator, is still related to the Voronoi cells. Its construction is carried out in the following way. First, the Voronoi cells for the points $\{x_i\}$ are calculated. Then, the point x is added to the set, and the new Voronoi cells are calculated. The cell corresponding to x in the new configuration overlaps

some parts (stolen areas) of cells originally owned by nearby points: these points, called *natural neighbors*, will be involved into the interpolation at x. Specifically, the value assigned to x will be the weighted average (see Eq. (30)) of values at the natural neighbors, with weights equal to the overlap areas (see Fig. 4).

The natural neighbor interpolator is linear, normalized, invariant upon translation, rotation, and scaling, exact, bounded, and weakly local. All these properties are easily verified except, perhaps, the exact property. Suppose that the point x moves toward one of the locations, say x_3. As x get closer and closer to x_3, the stolen area will almost be entirely inside the (old) cell corresponding to x_3. In the limit where x coincides with x_3, the stolen area will be the intersection of the cell of x_3 with the half-plane $\{\bar{x} \mid \langle x - x_3, \bar{x} \rangle > 0\}$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product.

Again, analytical calculations to obtain K for this smoothing technique are feasible only in dimension $n = 1$. In this case, surprisingly, the natural neighbor smoothing is totally equivalent to the Delaunay triangulation, and thus Eq. (62) holds.

For higher dimensions, we can only obtain numerical estimates for $K(|x - \bar{x}|)$. Figure 3 shows these estimates in dimension $n = 2$; in this case we obtain $\mathcal{N}_{\text{eff}} \simeq 3.41$.

3.4. Moving weights

One of the most common interpolating techniques used in Astronomy is based on a simplified form of Eq. (30) in which the ith weight depends only on x and on the ith location:

$$S\left(x; \{(x_i, y_i)\}\right) = \left[\sum_{i=1}^N w(x - x_i) y_i\right] \Big/ \left[\sum_{i=1}^N w(x - x_i)\right]. \tag{64}$$

The function $w(x - x_i)$, taken here to be non-negative, is usually chosen to have a peak at $x = x_i$ and to decrease to zero as $|x - x_i|$ increases. This way, the estimated value at x will be basically determined by the neighboring points. Often the weight function is isotropic, i.e. $w(x) = w(|x|)$. Commonly used weight functions include Gaussians, top-hats, and inverse distances of the form $w(x) = 1/|x|^\alpha$ with α fixed (this is precisely our toy-interpolator defined in Eq. (2)).

Some of the properties of the moving weight interpolator strongly depend on the weight function used. In general, this interpolator is linear, normalized, invariant upon translations, and bounded. Hence, the kernel associated to the smoothing satisfies $K(x - \bar{x}) \leq 1$ and is normalized to unit. The interpolator is invariant upon rotation if and only if w is isotropic. If w is homogeneous of degree g (as for example for inverse distances weights), i.e. if $w(kx) = k^g w(x)$, then the interpolator is invariant upon scaling . The interpolator is generally not exact; however, if $w(x)$ is bounded for $x \neq 0$ and goes to infinity for $x = 0$ (e.g., in case of inverse distance weights), then the interpolator is exact and thus we expect $K(0) = 1$. The interpolator is strongly local if and only if $w(x)$ has compact support, in which case K has also compact support; it is otherwise global. Finally, we observe that this interpolator is defined if just one point is inside the support of w. Hence, if w has compact

support of area a, we have $P_0(x) = \exp[-\rho a]$; in other cases $P_0(x)$ vanishes.

The statistical properties of this smoothing have been studied in detain in a separate paper (Lombardi & Schneider 2001). There we have shown that the kernel can be evaluated using the set of equations

$$Q(s) = \int_X \left[e^{-sw(x)} - 1 \right] dx, \tag{65}$$

$$K(x) = \frac{w(x)}{1 - P_0} \int_0^\infty e^{-w(x)s + \rho Q(s)} ds. \tag{66}$$

In the same paper we have also explicitly shown several of properties stated above for this interpolator. In addition, we have shown that the kernel K associated with this smoothing closely resemble the weight function w used, but is generally larger. For example, if we define the area $\mathcal{N}$ of w analogously to Eq. (48), we find $\mathcal{N}_{\text{eff}} \geq \mathcal{N}$. The kernel K however converges to w as $\rho \to \infty$.

3.5. Fixed weights

For some studies a non-normalized form of the smoothing (64) is still used:

$$S(x; \{(x_i, y_i)\}) = \sum_{i=1}^{N} w(x - x_i) y_i. \tag{67}$$

This smoothing has been employed, e.g., in some weak lensing studies (Kaiser & Squires 1993; Tyson & Fischer 1995; Luppino & Kaiser 1997; Fischer 1999).

The interpolation procedure (67) lacks several of the properties described above. In particular, it is only linear and invariant upon translation. If $w(x) = w(|x|)$ is isotropic, then the rotation invariance holds. The interpolator is *not* normalized (which is quite unusual for a smoothing technique); however, if the weight function w is normalized to 1, i.e.

$$\rho \int_X w(x) \, dx = 1, \tag{68}$$

then Eq. (28) still holds. The interpolator is not exact and not bounded; it is strongly local if w has compact support, otherwise it is global; it is always defined, so that $P_0(x) = P_0(x, x') = 0$.

A statistical analysis of this interpolator is straightforward. The result obtained is that the kernel K coincides with the weight function w. As expected, the kernel is not necessary bounded superiorly and is not necessary equal to unit at $x = 0$ (see Lombardi et al. 2002 for a detailed discussion of this smoothing technique applied to weak lensing data).

3.6. Other interpolation techniques

Clearly, a complete analysis of all smoothing methods currently used would be impossible here; on the other hand, the general classification technique described in Sect. 2 should allow one to characterize other interpolators. We want to mention however a couple of interesting smoothing techniques, namely the splines and the kriging.

Splines are an interesting and complex topic (see Press et al. 1992 for a short introduction; a complete discussion can be found in de Boor 1978). They find several applications in many different fields in astronomy, and can also be used for "cosmetic treatments". Splines come in several variants with slightly different properties. For example, some variants of splines are exact interpolators, while others are not. Another peculiarity is that often splines are *not* bounded interpolators. In order to avoid any possible confusion among the different variants of splines, we prefer here not to consider this smoothing technique in detail. We stress again, however, that a statistical analysis of any particular variant can be carried out without difficulties using the framework described in this paper.

Kriging is another interesting interpolation method. Although it is seldom used in Astronomy, it finds several applications in geophysics (see, however, Alfaro et al. 1991 for an example of astronomical application of kriging). A complete discussion of this interpolation method, which also has some variants, is beyond the scope of this paper. Here, we just observe that kriging is a non-linear method, and thus cannot profitably be analyzed with the techniques discussed in this article.

4. Conclusions

In this paper we have considered the statistical properties of interpolation techniques. The discussion has originally been kept at a high level of abstraction, and this has given us the ability to characterize smoothing methods in terms of simple properties. In particular, we have made a classification of interpolation techniques and we have derived several statistical results associated with each class of interpolators. A comparison of this analysis with a more technical one carried out in a separate paper for a specific smoothing method (Lombardi & Schneider 2001) clearly shows the advantages of using an abstract approach to the problem. In the second part of this paper we have applied the results obtained to several commonly used smoothing methods.

References

Adler, R. J. 1991, The geometry of random fields (Wiley & sons, New York)
Alfaro, E. J., Cabrera-Cano, J., & Delgado, A. J. 1991, ApJ, 378, 106
Bernardeau, F., & van de Weygaert, R. 1996, MNRAS, 279, 693
Cressie, N. A. C. 1993, Statistics for Spatial Data (Wiley & sons, New York)
de Boor, C. 1978, A Practical Guide to Splines (New York: Springer)
Fischer, P. 1999, AJ, 117, 2024
Kaiser, N., & Squires, G. 1993, ApJ, 404, 441
Lombardi, M., & Alves, J. 2001, A&A, 377, 1023
Lombardi, M., & Bertin, G. 1998, A&A, 335, 1
Lombardi, M., & Schneider, P. 2001, A&A, 373, 359

Lombardi, M., & Schneider, P. 2002, A&A, 392, 1153

Lombardi, M., Schneider, P., & Morales-Merlino, C. 2002, A&A, 382, 769

Luppino, G. A., & Kaiser, N. 1997, ApJ, 475, 20

O' Rouche, J. 1994, Computational Geometry in C (Princeton: Princeton Univ. Press)

Okabe, A., Boots, B., & Sugihara, K. 1992, Spatial tessellations: concepts and applications of Voronoi diagrams (John Wiley & Sons, New York)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C. The Art of Scientific Computing (Cambridge: Cambridge University Press)

Preston, C. 1976, Random fields (Springer-Verlag, Berlin)

Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169

Schaap, W. E., & van de Weygaert, R. 2000, A&A, 363, L29

Scott, E. L., Shane, C. D., & Swanson, M. D. 1954, ApJ, 119, 91

Sibson, R. 1981, in Interpreting multivariate data, ed. V. Barnett (John Wiley & Sons, Chichester), 21

Soneira, R. M., & Peebles, P. J. E. 1978, AJ, 83, 845

Tyson, J. A., & Fischer, P. 1995, ApJ, 446, L55

van de Weygaert, R. 1994, A&A, 283, 361

van Waerbeke, L. 2000, MNRAS, 313, 524

Watson, D. 1992, Contouring: a guide to the analysis and display of spatial data (Pergamon)