

Ajuste de datos por cuadrados mínimos

Least squares data fitting

Pedro Ripa†

CICESE

Recibido en febrero de 2001; aceptado en octubre de 2001

Resumen

Se desea representar, lo mejor posible, una serie de datos por medio de ciertas funciones con parámetros libres. “Lo mejor posible” significa que estos parámetros se eligen de manera que los residuos –la diferencia entre los datos y las funciones que se les ajustan– sean tan pequeños como se pueda. Nuestro objetivo no se limita a encontrar los parámetros del mejor ajuste, sino que también deseamos saber algo sobre sus incertidumbres, esto es, qué tan bien están determinados, dados tanto los errores de los datos originales como la imperfección del ajuste. Finalmente, suponiendo que utilizamos los parámetros del ajuste en el cálculo de otras variables, queremos también tener una estimación de las incertidumbres de estas últimas. Para poder avanzar, imaginamos propiedades básicas a las que llamamos “hipótesis”, y de ahí procedemos con rigor matemático. No es malo tener presente que las conclusiones a las que lleguemos dependen de las hipótesis hechas a lo largo del camino –incluyendo la idea de que del ajuste por cuadrados mínimos de esos datos, por estas funciones, se pueda extraer una información útil.

Palabras clave: Mínimos cuadrados, ajuste de datos.

Abstract

It is desired to represent, as good as possible, a series of data by means of certain functions with free parameters. “As good as possible” means that these parameters are chosen so that the residuals, the difference between data and fitting functions, be as small as it is feasible. Our objective is not limited to finding the parameters of the best fit, but we also wish to know something about their uncertainties, this is, how well they are determined, given the errors of the original data as well as the imperfection of the fitting. Finally, supposing that we use the parameters of the fit in the calculation of other variables, we also want to have an estimation of the uncertainties of the latter. In order to do that, we imagine basic properties, which we call “hypothesis”, and then proceed from there with mathematical rigor. It is not superfluous to remember that the conclusions at which we arrive depend on the hypotheses done throughout the way, including the idea that useful information can be extracted from a least squares fit, of those data by these functions.

Key words: Least squares, data fitting.

Comenzamos con un ejemplo muy sencillo

Supongamos que se tiene una serie de observaciones z_i ($i = 1, \dots, n$) de una misma variable. ¿Podemos extraer de todas ellas un solo número a al que le tengamos más confianza que a cada medida individual? ¿Cuál es la incertidumbre de ese número? Una forma de resolver este problema parte de escribir las observaciones como la suma del número buscado más residuos

$$z_i = a + r_i \tag{1}$$

y elegir a ese parámetro a como aquel que minimiza una cierta medida de los residuos, por ejemplo

$$\|r\|^2 = \sum_{i=1}^n w_i r_i^2 : \min \tag{2}$$

donde $w_i > 0$. En este caso sencillo, el parámetro buscado resulta ser un promedio ponderado de las observaciones:

Starting with a very simple example

Suppose one has a series of observations z_i ($i = 1, \dots, n$) of the same variable. Can we extract from all of them a single number a which we trust more than any single measurement? What is the uncertainty of that number? One way of solving this problem is writing the observations as the sum of the sought number plus residuals

and chose that parameter a as such that minimizes a certain measure of the residuals, for instance

where $w_i > 0$. In this simple case, the parameter wanted turns out to be a weighted average of the observations:

$$\frac{\partial \|r\|^2}{\partial a} = 0 \Rightarrow a = \frac{\sum_{i=1}^n w_i z_i}{\sum_{j=1}^n w_j} \quad (3)$$

El resultado de esta operación puede escribirse con tantas cifras como aguante la paciencia o se estire la tinta pero, ¿todas tienen sentido? Se necesita determinar la incertidumbre o error de a , a partir de alguna estimación de los errores de las observaciones y/o de que tan pequeños son los residuos.

Veamos el primer caso. Para poder avanzar, nos imaginamos a las observaciones z_i a nuestra disposición como la suma de un “valor verdadero” z_* (¡desconocido!) más unos “errores aleatorios” δz_i . Por aleatorios, queremos decir que, aunque tienen ciertos valores en las observaciones z_i , podrían haber tomado otros valores, aunque no de forma totalmente arbitraria: los valores que pueden tomar los errores δz_i —en series independientes de n observaciones cada una— están regidos por ciertas distribuciones de probabilidades. De estas distribuciones generalmente nos van a interesar sólo los dos primeros “momentos”, como se verá a continuación. En suma, imaginamos que los $\{z_i\}$ son un caso particular, una “realización”, de otros casos que podrían haberse obtenido en otras series de n observaciones cada una. En éste y otros ejemplos, vamos a indicar por $\langle \rangle$ el “valor esperado” o promedio estadístico sobre muchas realizaciones. No hay que confundir $\langle z_i \rangle$, para un cierto i , con el promedio (3). [Para aclarar este punto, consideremos que —en otros ejemplos— los z_i podrían ser observaciones de variables distintas, con unidades diferentes: en ese caso (3) no tendría sentido pero $\langle z_i \rangle$ sí].

La ecuación (3) indica que el parámetro a es una función de las z_i (en cada realización, o serie de observaciones). Luego, a es también una variable aleatoria: aunque tiene un valor determinado para las observaciones de que disponemos, podría haber tomado otros valores (si las observaciones hubieran tenido otros errores), no en forma totalmente arbitraria, sino también controlada por una cierta distribución de probabilidades.

La primera hipótesis que hacemos es que cada error δz_i (para un valor dado de i) puede tomar valores de diferentes signos y magnitudes, de forma tal que su valor esperado sea nulo: en las observaciones no hay *errores sistemáticos*. Esta hipótesis tiene una consecuencia importante sobre la distribución del parámetro a : su valor esperado es igual al valor verdadero de la variable z_* , es decir

$$\left. \begin{aligned} z_i &= z_* + \delta z_i \\ \langle \delta z_i \rangle &= 0 \end{aligned} \right\} \Rightarrow \left\{ \begin{aligned} a &= z_* + \delta a \\ \langle \delta a \rangle &= 0 \end{aligned} \right. \quad (4)$$

Esto implica que la relación (3) también se cumple entre los errores aleatorios de parámetro y datos, es decir

$$\delta a = \frac{\sum_{i=1}^n w_i \delta z_i}{\sum_{j=1}^n w_j} \quad (5)$$

The result of this operations can be written with as many digits as patience endures or as the ink stretches but, do they all make sense? It is necessary to determine the uncertainty of a , from some estimations of the observational errors and/or the smallness of the residuals.

Let us see the first case. In order to proceed, we imagine the observations z_i at our disposal as the sum of a “true value” z_* (unknown!) plus some “random errors” δz_i . By random we mean that, even though they have certain values in the observations z_i , they could have taken different values, but not in a totally arbitrary way: the values that the errors δz_i can take —in independent series with n observations each— are controlled by certain probability distributions. We will generally be interested in the first two “momenta” of those distributions, as shown below. In sum, we imagine that the $\{z_i\}$ are but a particular case, a “realization”, of other cases which could have been obtained in other series of n observations each. In this and other examples, we will indicate by $\langle \rangle$ the “expected value” or statistical average over many realizations. It is important not to confuse $\langle z_i \rangle$, for a certain i , with the average (3). [In order to clarify this point, consider that —in other examples— the z_i could be the observations of different variables, with different units: in such a case (3) would make no sense but $\langle z_i \rangle$ would].

Equation (3) shows that parameter a is a function of the z_i (in each realization, or series of observations). Then, a is also a random variable: even though it takes a certain value for the observations at our disposal, it could have taken other values (if the observations would have had other errors), non in arbitrary way, but also controlled by a probability distributions.

The first hypothesis we do is that each error δz_i (for a given i) may have different signs and magnitudes, such that its expected value vanishes; there are no systematic errors in the observations. This hypothesis has an important consequence on the distributions of parameter a : its expected value equals the true value of the variable z_* , namely

This implies that (3) also holds between the random errors of parameter and data, i.e.

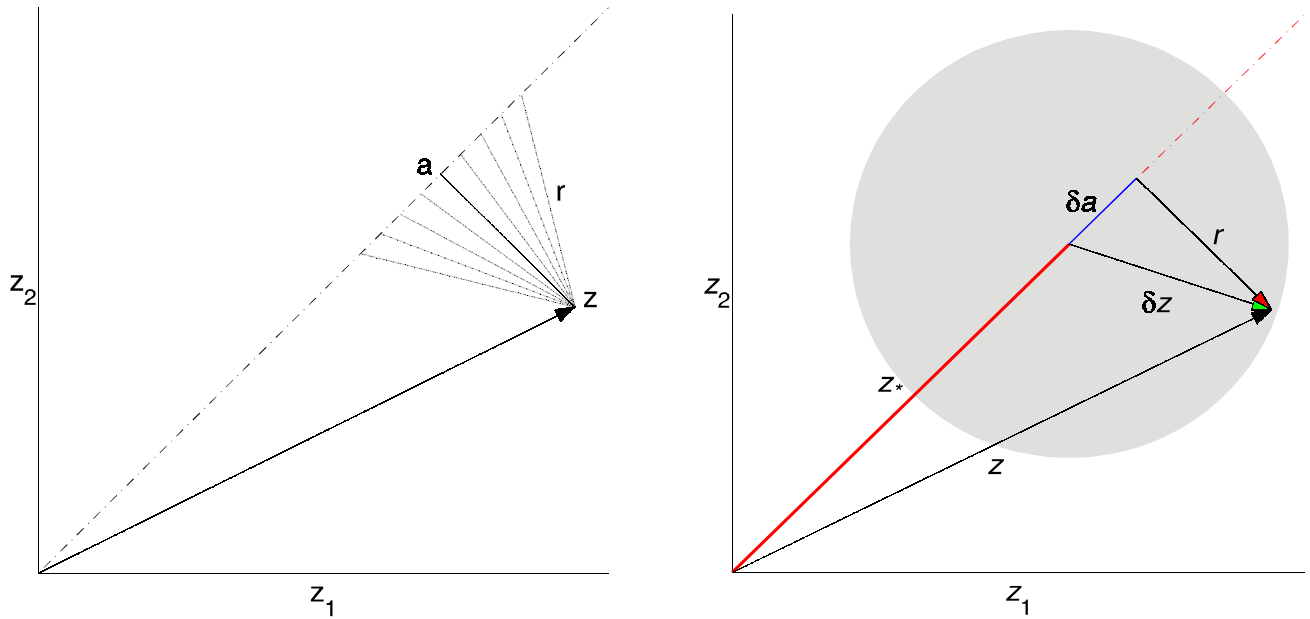


Figura 1. Promedio de dos números como un “ajuste por cuadrados mínimos”. Izquierda. El promedio a de dos números –las componentes el vector z – es el que minimiza el residuo r ; el residuo mínimo es perpendicular a la dirección del ajuste. Derecha. El círculo indica las observaciones z posibles; nótese que el error $\delta a = |a - z_*|$ del ajuste es menor que el error de la observación $\delta z = |z - z_*|$, donde z_* es el valor verdadero, salvo en el caso inusual de $r = 0$.

Figure 1. Average of two numbers as a “least squares fit”. Left. The average a of two numbers –the components of vector z – is that which minimizes the residue r ; the minimum residue is perpendicular to the fitting direction. Right. The circle shows the possible observations z ; note that the parameter error $\delta a = |a - z_*|$ is smaller than the observation error $\delta z = |z - z_*|$, where z_* is the true value, except in the unusual case $r = 0$.

El parámetro a es nuestra estimación de z_* , en base a una cantidad finita de observaciones $\{z_i\}$; es de esperarse que si el número n de observaciones independientes creciera indefinidamente, la diferencia $a - z_*$ o, más precisamente, la incertidumbre cuadrática media $\sqrt{\langle \delta a^2 \rangle}$, debería tender a cero.

Parameter a is the estimation of z_* based on a finite number of observations $\{z_i\}$; it is expected that as the number n of independent observations grows, the difference $a - z_*$ or, better, the root mean square uncertainty $\sqrt{\langle \delta a^2 \rangle}$, should tend to zero.

Uncertidumbre debida a los errores de los datos

Uncertainty due to data errors

Para llegar a una estimación de la incertidumbre $\langle \delta a^2 \rangle$ de a necesitamos hipótesis adicionales, sobre las incertidumbres de las observaciones z_i , concretamente, sobre los momentos de segundo orden de la distribución de los errores δz_i . Como las z_i entran en la estimación (3) de a , no es suficiente conocer el valor cuadrático medio de cada error $\langle \delta z_i^2 \rangle$, sino que también es necesario conocer el valor esperado del producto de errores diferentes, $\langle \delta z_i \delta z_j \rangle$ para $i \neq j$.

In order to estimate the uncertainty $\langle \delta a^2 \rangle$ of a we need additional hypotheses, on the uncertainties of the observations z_i , more precisely, on the second order momenta of the errors δz_i probability distribution. Since all the z_i enter in the estimation (3) of a , it is not enough to know the mean square error of each variable $\langle \delta z_i^2 \rangle$, it is also needed the expected value of the product of different errors $\langle \delta z_i \delta z_j \rangle$ for $i \neq j$.

El valor esperado $\langle \delta z_i \delta z_j \rangle$ es el promedio estadístico de $\delta z_i \delta z_j$ sobre un número infinito de realizaciones. Es algo que desconocemos, sólo tenemos una realización, una sola serie de n datos $\{z_i\}$. Por lo tanto, a falta de más información, vamos a suponer por el momento la situación más sencilla: los errores de cada observación z_i tienen igual valor cuadrático medio y no están correlacionados

The expected value $\langle \delta z_i \delta z_j \rangle$ is the statistical average of $\delta z_i \delta z_j$ over an infinite number of realizations. We don't know it, we only have one realization, only one series of n data $\{z_i\}$. Therefore, lacking more information we will assume the simplest situation, for the time being: the errors of all observations z_i have the same mean square and are not correlated

$$\langle \delta z_i \delta z_j \rangle = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}; \quad (6)$$

con ésto, podemos calcular fácilmente la incertidumbre de a : Elevando (5) al cuadrado y calculando el promedio estadístico resulta

$$\langle \delta a^2 \rangle = \frac{\sum_{i,k=1}^n w_i w_k \langle \delta z_i \delta z_k \rangle}{\left(\sum_{j=1}^n w_j\right)^2} = \frac{\sum_{i=1}^n w_i^2 \sigma^2}{\left(\sum_{j=1}^n w_j\right)^2}.$$

[La delta δ_{ij} “transforma” a la suma doble en el numerador en una suma sencilla; en general $\sum_{i,k=1}^n (\dots i \dots k) \delta_{ik} = \sum_{i=1}^n (\dots i \dots i)$]. Hasta ahora, no hemos dicho nada sobre los pesos w_i (salvo el hecho de que tienen que ser positivos). Parece atinado elegirlos de manera tal de *minimizar la incertidumbre del parámetro a* , recién calculada:

$$\frac{\partial \langle \delta a^2 \rangle}{\partial w_i} = 0 : \begin{cases} w_i = 1 \\ \langle \delta a^2 \rangle = \sigma^2 / n \end{cases} \quad (7)$$

Es decir, la incertidumbre mínima de a se obtiene para pesos w_i idénticos y es igual a la incertidumbre σ de cada observación, dividida por \sqrt{n} . ¡Esto es lo que hemos ganado al tomar todas las observaciones juntas! Nótese que $\sqrt{\langle \delta a^2 \rangle}$ tiende a cero conforme $n \rightarrow \infty$ (aunque lo hace lentamente, como $1/\sqrt{n}$) tal y como se mencionó al final de la subsección anterior.

No es esencial que los $w_i = 1$ para que $\langle \delta a^2 \rangle$ sea mínimo, sólo se necesita –en este caso– que los w_i sean iguales. La estimación óptima es la más ‘democrática’: todas y cada una de las observaciones tiene una contribución igual.

Uso de los cuadrados mínimos

Hasta el momento no hemos analizado los residuos o, más precisamente, su medida cuadrática $\|r\|^2$, definida en (2). Ésta es también una variable aleatoria ya que depende de los errores de las observaciones. En este ejemplo sencillo, de (1) y (4) es

$$\delta z_i = \delta a + r_i ;$$

elevando al cuadrado y utilizando $\sum_{i=1}^n w_i r_i = 0$, que es una consecuencia de (1) y (3), se encuentra que la medida de los residuos $\|r\|^2$ es menor que la de los errores $\|\delta z\|^2$ debido a que ha elegido el parámetro a para minimizar a la primera. Más concretamente,

$$\|r\|^2 = \|\delta z\|^2 - \|\delta a\|^2. \quad (8)$$

[La figura 1 ejemplifica cómo el promedio de dos números ($n = 2$) puede ser aquel que minimiza cuadrados $\sum_{i=1}^n (z_i - a)^2$ y que en el mínimo se cumple la relación pitagórica (8).]

with this, we can easily calculate the uncertainty of a : Squaring (5) and evaluating the statistical average it is

[δ_{ij} “transforms” the double sum in the numerator into a very simple single sum; in general $\sum_{i,k=1}^n (\dots i \dots k) \delta_{ik} = \sum_{i=1}^n (\dots i \dots i)$]. Up to now, we have said nothing about the weights w_i (except that they must be positive). It seems reasonable to choose them so that *the uncertainty of parameter a is minimized*:

Namely, the minimum uncertainty of a is obtained with identical weights w_i and is equal to the uncertainty σ of each observation divided by \sqrt{n} . This is what we gain by taking all the observations together! Note that $\sqrt{\langle \delta a^2 \rangle}$ tends to zero as $n \rightarrow \infty$ (albeit slowly, as $1/\sqrt{n}$) as mentioned at the end of the former subsection.

It is not essential that $w_i = 1$ for $\langle \delta a^2 \rangle$ to be minimum, it is only needed –in this case– that the w_i be equal. The optimum estimation is the more ‘democratic’: all observations have an equal contribution.

Use of the least squares

We haven’t analyzed so far the residuals or, more precisely, their measure $\|r\|^2$, defined in (2). This is also a random variable since it is a function of the observation errors. In this simple example, from (1) and (4) it is

squaring and using $\sum_{i=1}^n w_i r_i = 0$, which is a consequence of (1) and (3), it is found that the residuals measure $\|r\|^2$ is smaller than the error measure $\|\delta z\|^2$ because the parameter a has been chosen to minimize the former. More precisely,

[Figure 1 shows how the average of two numbers ($n = 2$) is that which minimizes the squares $\sum_{i=1}^n (z_i - a)^2$, and that at the minimum the Pythagorean relationship (8) holds true].

La ecuación (8), con δa de (5) y $w_i = 1$, expresa la relación entre $\sum_{i=1}^n r_i^2$ y los errores $\{\delta z_i\}$ de una realización en particular. La hipótesis (6) permite calcular el *valor esperado* $\sum_{i=1}^n \langle r_i^2 \rangle$ sobre todas las realizaciones. Concretamente es $\sum_{i=1}^n \langle \delta z_i^2 \rangle = n\sigma^2$ y $\sum_{i=1}^n \langle \delta a^2 \rangle = \sigma^2$, por lo que

$$\sum_{i=1}^n \langle r_i^2 \rangle = (n-1)\sigma^2. \quad (9)$$

Esta ecuación es útil cuando no se conoce el “error por punto” σ . En ese caso, es común elegir σ de manera que (9) se cumpla para el valor particular de $\|r\|^2$ calculado, que no es lo mismo que el valor esperado $\langle \|r\|^2 \rangle$. De esa forma

$$\hat{\sigma}^2 : \begin{cases} w_i = 1 \\ \sigma^2 = \frac{\sum_{i=1}^n r_i^2}{n-1} \end{cases},$$

con lo que (7) da

$$\langle \delta a^2 \rangle = \frac{\sum_{i=1}^n r_i^2}{n(n-1)}. \quad (10)$$

Con más datos, o sea para un valor mayor de n , $\sum_{i=1}^n r_i^2$ aumenta (aproximadamente proporcional a $n-1$) y por lo tanto disminuye la incertidumbre $\sqrt{\langle \delta a^2 \rangle}$ (aproximadamente como $1/\sqrt{n}$), pero esto requiere de que las observaciones sean estadísticamente independientes $\langle \delta z_i \delta z_j \rangle = 0$ para $i \neq j$. Si se aumentara el número de observaciones “sobremuestreando”, o sea tomando datos con errores muy correlacionados, entonces lógicamente $\sqrt{\langle \delta a^2 \rangle}$ no debería disminuir sensiblemente.

Por otra parte, si sí se conoce σ *a priori* (o hay razones independientes del ajuste para suponerle algún valor) entonces (9) puede utilizarse para comparar la medida de los residuos $\|r\|^2$ obtenida de los datos $\{z_i\}$ con su valor esperado o promedio estadístico $(n-1)\sigma^2$. Si $\|r\|^2$ fuera mucho menor o mucho mayor que el valor esperado $(n-1)\sigma^2$, entonces estos $\{z_i\}$ tendrían errores muy inusuales ... o σ no estaría bien estimado. En el primer caso (ajuste inusualmente malo) es conveniente redefinir σ , aumentándolo para que se cumpla $\sigma^2 = \|r\|^2 / (n-1)$, lo que resulta en la incertidumbre (10) para el parámetro a . ¿Qué tanto tiene que diferir $\|r\|^2 / \sigma^2$ de $n-1$ para que el ajuste sea sospechoso? En general se espera una distribución de los valores de los residuos tal que la mayoría de los casos caigan en la banda¹

$$\frac{\sum_{i=1}^n r_i^2}{\sigma^2} = (n-1) \pm \sqrt{n-1}. \quad (11)$$

Equation (8), with δa from (5) and $w_i = 1$, expresses the relationship between $\sum_{i=1}^n r_i^2$ and the errors $\{\delta z_i\}$ in a particular realization. Hypothesis (6) allows the calculation of the expected value $\sum_{i=1}^n \langle r_i^2 \rangle$ over all realizations. Namely, it is $\sum_{i=1}^n \langle \delta z_i^2 \rangle = n\sigma^2$ and $\sum_{i=1}^n \langle \delta a^2 \rangle = \sigma^2$, and therefore:

This is a useful equation when the “error per point” σ is not known. In such a case, it is common to choose σ such that (9) is fulfilled for the particular value of $\|r\|^2$ calculated, which is not the same as the expected value $\langle \|r\|^2 \rangle$. Thus

and therefore (7) gives

With more data, increasing n , it is expected that $\sum_{i=1}^n r_i^2$ will grow (approximately as $n-1$) and therefore the uncertainty $\sqrt{\langle \delta a^2 \rangle}$ will decrease (approximately as $1/\sqrt{n}$), but this requires that the observations be statistically independent $\langle \delta z_i \delta z_j \rangle = 0$ for $i \neq j$. If the number of observations is increased by “oversampling”, that is, taking data with highly correlated errors, then it is not expected that $\sqrt{\langle \delta a^2 \rangle}$ will diminish sensibly.

On the other hand, if σ is known *a priori* (or there are reasons independent from the fitting to assume a certain value for this error) then (9) can be used to compare the residuals measure $\|r\|^2$ obtained from the data $\{z_i\}$ with its expected value or statistical average $(n-1)\sigma^2$. If $\|r\|^2$ were much larger or smaller than the expected value $(n-1)\sigma^2$, then these $\{z_i\}$ would have very unusual errors ... or σ would not be well estimated. In the first case (unusually bad fit) it is convenient to redefine σ , increasing it so $\sigma^2 = \|r\|^2 / (n-1)$ is satisfied, which results in the uncertainty (10) for parameter a . How much does $\|r\|^2 / \sigma^2$ have to differ from $n-1$ in order for the fit to be suspicious? It is expected, in general, that in most cases the residuals fall in the band¹

Complicando un poco el ejemplo

En vez de (6), vamos ahora a suponer que los errores δz_i tienen diferentes valores cuadráticos medios $\langle \delta z_i^2 \rangle$, aunque siguen sin estar estadísticamente correlacionados, es decir,

$$\langle \delta z_i \delta z_j \rangle = \epsilon_i^2 \delta_{ij} = \begin{cases} \epsilon_i^2 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad (12)$$

Este caso puede darse cuando las observaciones $\{z_i\}$ están hechas independientemente y con diversos instrumentos, de precisión diferente. De (5) es ahora

$$\langle \delta a^2 \rangle = \frac{\sum_{i=1}^n w_i^2 \epsilon_i^2}{(\sum_{j=1}^n w_j)^2}$$

Minimizando esta expresión con respecto a los pesos, se obtiene

$$\frac{\partial \langle \delta a^2 \rangle}{\partial w_i} = 0 : \begin{cases} w_i = \epsilon_i^{-2} \\ \langle \delta a^2 \rangle = 1 / \sum_{j=1}^n \epsilon_j^{-2} \end{cases} \quad (13)$$

en vez de (7). Los pesos óptimos son ahora inversamente proporcionales al cuadrado de los errores de las observaciones. Los datos con mayor incertidumbre ϵ_i “pesan menos” en (3), aunque sí son tomados en cuenta. Con los pesos $w_i = \epsilon_i^{-2}$ se puede demostrar que (ver por ejemplo la nota 7, para el caso más general)

$$\langle \|r\|^2 \rangle = \sum_{i=1}^n \frac{\langle r_i^2 \rangle}{\epsilon_i^2} = n - 1 \quad (14)$$

en vez de (9). La distribución esperada de los residuos es otra vez la de (11), $\|r\|^2 = (n - 1) \pm \sqrt{n - 1}$; si el ajuste da un valor de $\|r\|^2$ demasiado grande, respecto de esta predicción, es conveniente aumentar los ϵ_i proporcionalmente (con la cual $\sqrt{\langle \delta a^2 \rangle}$ aumenta en la misma proporción) de manera que se cumpla $\|r\|^2 = (n - 1)$. [Se puede verificar que las fórmulas del ejemplo anterior, correspondientes a la distribución de errores de las observaciones (6), son un caso particular de las de éste, reemplazando los ϵ_i por σ].

¿Cómo se representa gráficamente este caso? En la figura 1, la zona de incertidumbre no se vería como un círculo sino como una elipse cuyos ejes principales están en las direcciones horizontal y vertical, y el vector residuo no se vería perpendicular a la dirección de ajuste (la recta al origen de 45°). Sin embargo, cambiando de espacio de graficado, mediante la transformación de escala $(z_1, z_2) \rightarrow (\sqrt{w_1}z_1, \sqrt{w_2}z_2)$, la nube de incertidumbre se vería otra vez como un círculo y el vector residuo sería ortogonal a la dirección de ajuste, sólo que ésta no estaría a 45°, sino que tendría un ángulo $\tan^{-1} \sqrt{w_2/w_1}$ con la horizontal.

Making the example a bit more complicated

Instead of (6), we will now assume that the errors δz_i have different mean square expected values $\langle \delta z_i^2 \rangle$, although they are still statistically uncorrelated, i.e.

This is for instance the case when the observations $\{z_i\}$ are made independently with diverse instruments that have different precision. From (5) it is now

Minimizing this expression with respect to the weights, it is obtained

instead of (7). The optimal weights are now inversely proportional to the observations squared errors. Data with more uncertainty ϵ_i “weight less” in (3), even though they are all taken into account. With the weights $w_i = \epsilon_i^{-2}$ it can be shown (see for instance note 7, for the most general case) that

instead of (9). The expected distribution of residuals is again (11), $\|r\|^2 = (n - 1) \pm \sqrt{n - 1}$; if the fitting gives a too large value of $\|r\|^2$, with respect to this prediction, it is convenient to increase the ϵ_i proportionally (and so $\sqrt{\langle \delta a^2 \rangle}$ increases in the same proportion) so that $\|r\|^2 = (n - 1)$ is fulfilled. [It can be verified that the formulae from last example, wich correspond to the observation error distribution (6), are but a particular case of those from this example, substituting the ϵ_i by σ].

How is this case represented graphically? The uncertainty zone in figure 1 would not be a circle

but and ellipse, whit principal axes in the horizontal and vertical directions, and the residual vector would not be seen perpendicular to the fitting direction (the straight line at 45° through the origin). However, changing the plotting space by means of the scale transformation $(z_1, z_2) \rightarrow (\sqrt{w_1}z_1, \sqrt{w_2}z_2)$, the uncertainty cloud would be seen again as a circle and the residual vector would be orthogonal to the fitting direction, except that this would not be at 45°, but it would have an angle $\tan^{-1} \sqrt{w_2/w_1}$ with the horizontal direction.

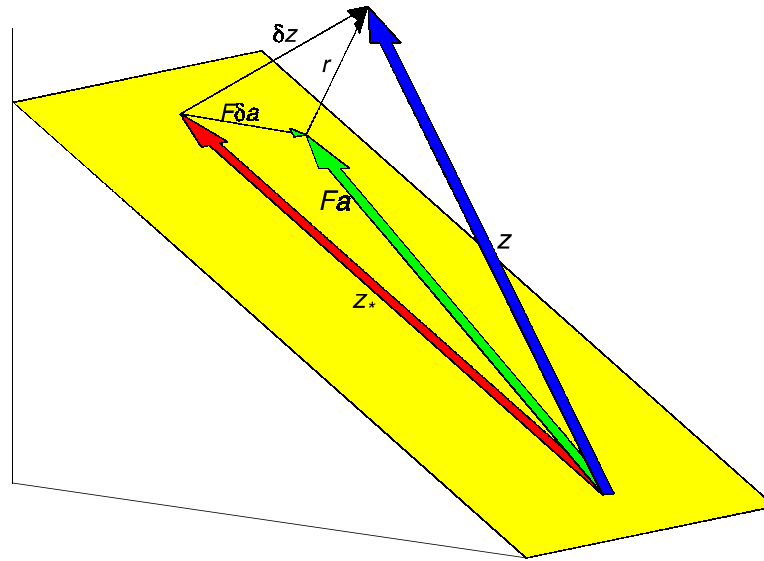


Figura 2. Ajuste de los datos z_i ($i = 1, \dots, n$) a funciones $Fa = \sum_{\mu=1}^m F_{\mu} a_{\mu}$; el ajuste óptimo es tal que el residuo $r = z - Fa$ es normal al espacio de ajustes posibles $\{z | z = Fa, a \in \mathbb{R}^m\} \in \mathbb{R}^n$, aquí representado por el plano sombreado. En esta ilustración es $n = 3$ y $m = 2$, pero la propiedad (18) de ortogonalidad entre Fa y r es general. El vector de datos con errores z “se sale” de ese espacio, pero se supone que el valor verdadero z_* está en él, es decir, que $z_* = Fa_*$ para un cierto a_* .

Figure 2. Fitting of data z_i ($i = 1, \dots, n$) to the functions $Fa = \sum_{\mu=1}^m F_{\mu} a_{\mu}$; the optimum fit is such that the residual $r = z - Fa$ is normal to the space of possible fits $\{z | z = Fa, a \in \mathbb{R}^m\} \in \mathbb{R}^n$, here represented by the shaded plane. In this illustration it is $n = 3$ and $m = 2$, but the orthogonality property (18) between Fa and r is general. The data with errors vector z “goes out” from that space, but is assumed that the true value z_* is on it, namely, it is $z_* = Fa_*$ for a certain a_* .

Propagación de errores

Supongamos que se desea calcular una función f de la variable z . El problema es que no conocemos el valor verdadero z_* de ésta, sino que sólo tenemos la estimación

$$z_* = a \pm \sqrt{\langle \delta a^2 \rangle},$$

expresada aquí en la forma habitual de “valor estimado \pm incertidumbre”. El método clásico de “propagación de errores” –en este caso: de la incertidumbre de a a la incertidumbre de f – se basa en linealizar las relaciones entre variables, elevarlas al cuadrado y calcular el valor esperado. Es decir, de $z_* = a + \delta a$ pasamos a $f(z_*) = f(a) - f'(a) \delta a + \dots$. Expresando esto como una variable aleatoria $f = f_* + \delta f$, estimamos su incertidumbre como $\langle \delta f^2 \rangle = f'(a)^2 \langle \delta a^2 \rangle$, es decir, se calcula el error cuadrático medio como el valor esperado de la contribución finita más baja. El resultado final puede expresarse en la forma

$$f(z_*) = f(a) \pm \sqrt{f'(a)^2 \langle \delta a^2 \rangle}.$$

Por ejemplo, si se desea calcular el área f de un cuadrado, en base a n medidas independientes z_i de su lado, entonces es

Error propagation

Assume that we wish to calculate a function f of variable z . The problem is that we don't know its true value z_* , we only have the estimation

expressed here in the usual “estimated value \pm uncertainty” form. The classical method of “error propagation” –in this case: from the uncertainty of a to the uncertainty of f – is based in linearizing the relationship between variables, squaring and calculating the expected value. That is, from $z_* = a + \delta a$ we get $f(z_*) = f(a) - f'(a) \delta a + \dots$. Writing this as a random variable $f = f_* + \delta f$, we estimate its uncertainty as $\langle \delta f^2 \rangle = f'(a)^2 \langle \delta a^2 \rangle$, namely, the mean square error is evaluated from the expected value of the lowest finite contribution. The final result can be expressed as

For instance, if the area f of a square is to be calculated from n independent measurements z_i of its side length, then is

$f = a^2 \pm 4a\sqrt{\langle \delta a^2 \rangle}$, o sea $f = \bar{z}^2 \pm 4\bar{z}\sqrt{\langle (z_i - \bar{z}^2)/(n-1) \rangle}$ donde la barra indica promedio aritmético sobre los datos. Más adelante se presentarán otros ejemplos, que son más interesantes por tener más parámetros.

$f = a^2 \pm 4a\sqrt{\langle \delta a^2 \rangle}$, i.e. $f = \bar{z}^2 \pm 4\bar{z}\sqrt{\langle (z_i - \bar{z}^2)/(n-1) \rangle}$ where the bar indicates arithmetic mean over the data. Other examples are presented below, which are more interesting on account of having more parameters.

El problema lineal más general

The more general linear problem

Vamos ahora a encarar el caso más general que deseamos ver aquí, con más de un parámetro, pero limitándonos al tipo de problema más sencillo, en el que las funciones de ajuste son una combinación lineal de los parámetros. Concretamente, se plantea ajustar n datos z_i (los que pueden o no ser las observaciones de variables distintas), mediante una expresión lineal con $m (\leq n)$ parámetros libres a_μ , es decir

We will now address a more general case, with more than one parameter, but restricting ourselves to the simplest problem, one in which the fitting functions are linear combinations of the parameters. More precisely, it is desired to fit n data z_i (which might be or not observations of different variables), by means of a linear expression with $m (\leq n)$ free parameters a_μ , i.e.

$$\underbrace{z}_{n \times 1} = \underbrace{F}_{n \times m} \underbrace{a}_{m \times 1} + \underbrace{r}_{n \times 1}, \tag{15}$$

donde se ha utilizado la notación matricial; las variables debajo de cada matriz indican su dimensión (número de filas \times número de columnas)². Los coeficientes a se determinan minimizando una medida de los residuos (los “cuadrados” del título):

where we have used matrix notation; the variables below each matrix indicate its dimension (number of rows \times number of columns)². The coefficients a are determined minimizing a residuals measure (the “squares” in the title):

$$\|r\|^2 = r^T W r : \min \tag{16}$$

donde W es una matriz “definida positiva” de $n \times n$ (o sea, $r^T W r > 0$ para todo $r \neq 0$). La solución es³

where W is a $n \times n$ “positive definite” matrix (namely, $r^T W r > 0$ for every $r \neq 0$). The solution is³

$$\frac{\delta \|r\|^2}{\delta a} = 0 \Rightarrow \begin{cases} a = Sz \\ S := MF^T W \\ M := (F^T W F)^{-1} \end{cases} \tag{17}$$

Para que exista la matriz M , de $m \times m$, no debe haber redundancias en las F 's, es decir, no debe existir ningún a_0 tal que $Fa_0 = 0$: en ese caso M es definida positiva. (Si $m \geq n$ y las F 's son linealmente independientes, el ajuste es perfecto, los residuos son idénticamente nulos). Más adelante discutiremos los casos no muy bien planteados, en los que aunque estrictamente $(F^T W F)$ tiene inversa, algunos parámetros a_μ están fuertemente correlacionados.

In order for the $m \times m$ matrix M to exist, there must be no redundancies among the F 's, namely, there must be no a_0 such that $Fa_0 = 0$: in this case M is positive definite. (If $m \geq n$ and the F 's are linearly independent, the fit is perfect, the residuals vanish identically). We will discuss below not-too-well-posed-cases, for which even though $(F^T W F)$ strictly has an inverse, some of the parameters a_μ are strongly correlated.

Espacios ortogonales

Orthogonal spaces

Es interesante ver que el ajuste óptimo Fa y los residuos r correspondientes son las componentes de los datos en *espacios ortogonales*: Definiendo las matrices P y Q mediante

It is interesting to see that the optimum fit Fa and the corresponding residuals r are the components of the data in *orthogonal spaces*: Defining the matrices P and Q by

$$\begin{pmatrix} Fa \\ r \end{pmatrix} = \begin{pmatrix} FS \\ I_n - FS \end{pmatrix} z = \begin{pmatrix} P \\ Q \end{pmatrix} z$$

se puede demostrar⁴ que son proyectores ortogonales, es decir, satisfacen

it can be shown⁴ that they are orthogonal projectors, that is, they satisfy

$$P^2 = P, Q^2 = Q, PQ = QP = 0_n;$$

P proyecta en el subespacio de todos los ajustes posibles, mientras que Q lo hace en el subespacio de los residuos, la parte “no ajustable” de los datos (véase figura 2). El teorema de Pitágoras correspondiente es $\|z\|^2 = \|Fa\|^2 + \|r\|^2$ o, despejando la medida de los residuos y escribiendo la relación más explícitamente,⁵

P projects into the subspace of all the possible fits, while Q does it in the subspace of the residuals, the “non-adjustable” part of the data (see figure 2). The corresponding Pythagorean theorem is $\|z\|^2 = \|Fa\|^2 + \|r\|^2$ or, solving for the residuals measure and writing in matrix notation,⁵

$$r^T W r = z^T W z - a^T (F^T W F) a, \quad (18)$$

donde se ve que la noción de ortogonalidad depende de la métrica W utilizada. Será interesante ver bajo qué condiciones esta descomposición es también válida para los errores, en una generalización de (8).

where it is seen that the notion of orthogonality depends on the metric W used. It will be interesting to see under which conditions this decomposition is also valid for the errors, in a generalization of equation (8).

Incertidumbre debida a errores de los datos

Uncertainty due to the data errors

Aunque (17) es la solución del problema de cuadrados mínimos, nos interesa ir más lejos: estimar la incertidumbre de los coeficientes a y de predicciones hechas en base a ellos. Como antes, en primer lugar vamos a suponer que los datos no tienen errores sistemáticos, luego

Although (17) it is the solution of the least squares problem, we want to go farther: to estimate the uncertainties of the coefficients a and of predictions made from them. As before, we will assume that the data do not have systematic errors, then

$$\left. \begin{array}{l} z = z_* + \delta z \\ \langle \delta z \rangle = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} a = a_* + \delta a \\ \langle \delta a \rangle = 0 \end{array} \right. \quad (19)$$

donde de (17) es

where from (17) it is

$$a_* := \langle a \rangle = S z_*, \delta a = S \delta z \quad (20)$$

Para caracterizar estadísticamente los errores δz es necesario, en general, conocer $\langle \delta z_i \delta z_j \rangle$ ($i, j = 1, \dots, n$), que son las componentes de la matriz $\langle \delta z \delta z^T \rangle =: C_z$. Para calcular las incertidumbres C_a de los parámetros a , en función de las incertidumbres C_z de los datos z , o sea

In order to characterize statistically the errors δz it is necessary, in general, to know $\langle \delta z_i \delta z_j \rangle$ ($i, j = 1, \dots, n$), which are the components of the matrix $\langle \delta z \delta z^T \rangle =: C_z$. To calculate the uncertainties C_a of the parameters a , from the data z uncertainties C_z , that is

$$C_z = \langle \delta z \delta z^T \rangle \xrightarrow{?} C_a = \langle \delta a \delta a^T \rangle$$

se aplica la técnica de propagación de errores a la segunda relación en (20), lo que resulta en⁶

the error propagation technique is applied to the second relationship in (20), which gives⁶

$$C_a = S C_z S^T$$

donde es bueno recordar que S depende de la matriz de pesos W que se ha elegido para hacer el ajuste. Si elegimos a W de manera que los parámetros a_μ estén determinados lo mejor

where it is good to remember that S depends on the weights' matrix W chosen for the fitting. If we choose W so that the parameters a_μ are determined as best as possible, then the form

posible, entonces la forma de C_a se simplifica bastante. Concretamente, en el apéndice A se demuestra que

$$\frac{\delta C_a}{\delta W} = 0 : \begin{cases} W = C_z^{-1} \\ C_a = M \end{cases} \quad (21)$$

Estos “pesos óptimos” W son la generalización de los casos de la forma (12), donde datos con mayor error pesan menos. [De hecho, cambiando de coordenadas diagonalizando C_z , se puede ver que (21) corresponde exactamente a (13). Sin embargo, no siempre es útil hacer tal cambio de coordenadas, v.g., cuando las z_i tienen un significado especial].

Uso de los cuadrados mínimos

Ahora queremos emprender el paso siguiente: utilizar los residuos para cuantificar la “bondad del ajuste”, si se conoce C_z , o para estimar estos errores, si no se los conoce. Lamentablemente –y a diferencia del primer ejemplo sencillo estudiado en (4)– las relaciones (20), entre los valores esperados y los errores de z y a , no garantizan que los residuos se deban exclusivamente a los errores, “ $r = \delta z - F\delta a$ ”, ya que el valor verdadero de los datos z_* , podría tener una proyección finita en el espacio de los residuos, $r_* := Qz_* \neq 0$ (el vector z_* “se sale” del plano de ajustes, en el ejemplo de la figura 2), es decir

$$\text{si } r_* := \langle r \rangle \neq 0 : \begin{cases} \delta z = F\delta a + \delta r \\ r = r_* + \delta r \end{cases}$$

De la primera expresión se puede fácilmente deducir un teorema tipo Pitágoras, similar a (18): usando $F\delta a = P\delta z$ y $\delta r = Q\delta z$, tomando sus medidas cuadráticas y calculando el promedio estadístico se encuentra⁷

$$\begin{aligned} \langle \|\delta r\|^2 \rangle &= \langle \|\delta z\|^2 \rangle - \langle \|F\delta a\|^2 \rangle \\ &= n - m, \end{aligned}$$

donde hemos supuesto que se están usando los pesos óptimos (21). De la segunda ecuación resulta

$$\langle \|r\|^2 \rangle = \|r_*\|^2 + \langle \|\delta r\|^2 \rangle,$$

ya que los términos cruzados $r_*^T W \delta r + \delta r^T W r_*$, tienen valor esperado nulo. Claramente el primer término, desconocido, no permite comparar a $\|r\|^2$ con $(n - m)$, como hicimos con el primer ejemplo sencillo con que comenzamos.

Vamos, por lo tanto, a hacer la hipótesis –generalmente muy fuerte– de que *el valor verdadero de las observaciones z_**

of C_a is simplified a lot. Concretely, in the appendix A it is demonstrated that

These “optimum weights” W are the generalization of those from (12), where the data with more error weighed less. [In fact, changing coordinates diagonalizing C_z , one can see that (21) corresponds exactly to (13). However, it is not always useful to make such a change of coordinates, for example when the z_i have a special meaning].

Use of the least squares

Now we want to undertake the following step: to use the residuals to quantify the “goodness of the fit”, if one knows C_z , or to estimate these errors, if they are not know. Regrettably –unlike the case of the simple example studied in (4)– the relationships (20), between the expected values of z and a and their errors, do not guarantee that the residuals are owed exclusively to the errors, “ $r = \delta z - F\delta a$ ”, since the data true value z_* might have a finite projection into the residuals space, $r_* := Qz_* \neq 0$ (vector z_* “goes out” of the fitting plane, in the example of the figure 2), namely,

A Pythagorean theorem similar to (18) can be easily deduced from the first expression, using $F\delta a = P\delta z$ and $\delta r = Q\delta z$ and taking the statistical average of their measures it is found⁷

where we have assumed that the optimum weights (21) are used. From the second expression it further follows

since the crossed terms, $r_*^T W \delta r + \delta r^T W r_*$, have a vanishing expected value. Clearly the first term, unknown, does not allow the comparison of $\|r\|^2$ with $(n - m)$, as we made in the first simple example we began with.

We are going, therefore, to make the hypothesis –generally a very strong one– that *the observations true value z_* has a null*

tiene una proyección nula en el espacio de los residuos, $Qz_* = 0$. Dicho de otra forma, supondremos que z_* puede ser exactamente expresado como una de las funciones de ajuste, para ciertos coeficientes a_* . Bajo esta hipótesis

projection in the space of the residuals, $Qz_* = 0$. In other words, we will assume that z_* can be exactly expressed as one of the fitting functions, for certain coefficients a_* . Under this hypothesis

$$\boxed{Qz_* = 0 \Leftrightarrow \exists a_*: z_* = F a_*} \Rightarrow \langle \|r\|^2 \rangle = n - m \quad (22)$$

En general se espera

In general, it is expected

$$\|r\|^2 = (n - m) \pm \sqrt{n - m},$$

como antes. Si, por ejemplo, no se conoce C_z , los z_i tienen todas las mismas unidades y *a priori* se espera que tengan errores equivalentes, se puede hacer

as before. If, for example, C_z is not known, the z_i have all the same units and *a priori* they are expected to have equivalent errors, one can make

$$iC_z? : \begin{cases} C_z = \sigma^2 I_n \\ \sigma^2 = \frac{r^T r}{n - m} \end{cases} \quad (23)$$

Es importante insistir en que esta estimación depende de la hipótesis (22), la cual expresa que el valor verdadero de los datos (sin los errores aleatorios) tiene un ajuste perfecto. Por ejemplo, consideremos el caso de los promedios del nivel del mar en un puerto, para cada mes del año (este problema es discutido en más detalle más adelante en el ejemplo de la variación estacional. Los datos son doce números z_i , que escribimos como

It is important to insist in that this estimate depends on hypothesis (22), which expresses that the true value of the data (without the random errors) has a perfect fit. For instance, let us consider the case of the averages of the sea level at a port, for every month of the year (this problem will be considered in greater detail in the seasonal variation example. The data are twelve numbers z_i , which we write as

$$z_i = z_*(t_i) + \delta z_i.$$

Normalmente desconocemos tanto la dependencia temporal del valor verdadero de los datos $z_*(t)$, sin errores, como su estadística $\langle \delta z_i \delta z_j^T \rangle$. Este segundo aspecto será analizado más adelante. Con respecto al primero, podemos hacer la hipótesis de que es exactamente la senoide

We usually ignore both the time dependence $z_*(t)$ of the data true value, without errors, as well as the statistics $\langle \delta z_i \delta z_j^T \rangle$ of the latter. This second aspect will be analyzed later on. With respect to the first one, we can make the hypothesis that it is exactly the sinusoid

$$H_1 : z_*(t) = a_{*1} + a_{*2} \cos \omega t + a_{*3} \sin \omega t,$$

donde ω es un ciclo por año y $a_{*\mu}$ es el valor verdadero del parámetro $a_{*\mu}$, o la hipótesis alternativa de que es la suma de dos senoideas

where ω is one cycle per year and $a_{*\mu}$ is the true value of parameter $a_{*\mu}$, or the alternative hypothesis of that it is the sum of two sinusoids

$$H_2 : z_*(t) = a_{*1} + a_{*2} \cos \omega t + a_{*3} \sin \omega t + a_{*4} \cos 2\omega t + a_{*5} \sin 2\omega t.$$

Lo que es importante destacar en este momento es que las hipótesis H_1 y H_2 son incompatibles (y, probablemente, igual arbitrarias). Debemos elegir entre una y otra, para estimar los parámetros $a_{*\mu}$ y sus incertidumbres, y es bueno que quede claro que no es más que una hipótesis, quizás injustificada, pero necesaria para poder seguir adelante.

At this time it is important to point out that hypotheses H_1 and H_2 are incompatible (and, probably, equally arbitrary). We must choose one or the other, in order to estimate the parameters $a_{*\mu}$ and their uncertainties, and it is good to make clear that it is not more than a hypothesis, maybe unjustified, but necessary to be able to continue ahead.

Propagación de errores

Supongamos que se desea calcular una función f de las variables a . El problema es que no conocemos el valor verdadero a_* de éstas, sino que de

$$a = a_* + \delta a$$

sólo tenemos las estimaciones

$$a_* = Sz, \langle \delta a \delta a^T \rangle = M.$$

Linealizando $f(a) = f(a_*) + \nabla f(a_*) \delta a + \dots$, donde $\nabla f(a_*)$ es el vector fila (matriz de $1 \times m$) que contiene las derivadas parciales $\partial f(a) / \partial a_i$ en $a = a_*$, y expresando a f como una variable aleatoria $f = \bar{f} + \delta f$, calculamos su incertidumbre a partir de la contribución más baja del cuadrado, $\delta f^2 = (\nabla f)^T \delta a \delta a^T (\nabla f)^T$, cuyo valor esperado es

$$\langle \delta f^2 \rangle = (\nabla f) M (\nabla f)^T.$$

A título de ejercicio, demostrar que si, por ejemplo,

$$f = v_1 a_1 + v_2 a_2,$$

donde v_1 y v_2 son número conocidos, entonces

$$\langle \delta f^2 \rangle = v_1^2 \langle \delta a_1^2 \rangle + v_2^2 \langle \delta a_2^2 \rangle + 2v_1 v_2 \langle \delta a_1 \delta a_2 \rangle.$$

Por otra parte, si

$$f = a_1^{v_1} + a_2^{v_2},$$

entonces

$$\frac{\langle \delta f^2 \rangle}{f^2} = v_1^2 \frac{\langle \delta a_1^2 \rangle}{a_1^2} + v_2^2 \frac{\langle \delta a_2^2 \rangle}{a_2^2} + 2v_1 v_2 \frac{\langle \delta a_1 \delta a_2 \rangle}{a_1 a_2}.$$

Un último problema es el siguiente: Se desea hacer el ajuste

$$z_i = a_1 + a_2 \cos \omega t_i + a_3 \sin \omega t_i + r_i.$$

Si se supone que los errores de las observaciones a diferentes tiempos no están correlacionados y tienen igual valor cuadrático medio, $\langle \delta z_i \delta z_j \rangle = \sigma^2 \delta_{ij}$: (1) mostrar que el mejor ajuste es el que minimiza $r^T r$ (es decir, con pesos iguales); (2) ¿qué condición deben cumplir los tiempos de muestreo $\{t_i\}$ de manera que

$$\langle \delta a_2^2 \rangle = \langle \delta a_3^2 \rangle \text{ y } \langle \delta a_2 \delta a_3 \rangle = 0?$$

Error propagation

Let us suppose that we want to calculate a function f of the variables a . The problem is that we don't know their true value a_* , but rather in

$$a = a_* + \delta a$$

we only have the estimates

$$a_* = Sz, \langle \delta a \delta a^T \rangle = M.$$

Linearizing $f(a) = f(a_*) + \nabla f(a_*) \delta a + \dots$, where $\nabla f(a_*)$ it is the row vector ($1 \times m$ matrix) that contains the partial derivatives $\partial f(a) / \partial a_i$ at $a = a_*$, and expressing f as a random variable $f = \bar{f} + \delta f$, we calculate its uncertainty from the square's lowest contribution, $\delta f^2 = (\nabla f)^T \delta a \delta a^T (\nabla f)^T$, whose expected value is

$$\langle \delta f^2 \rangle = (\nabla f) M (\nabla f)^T.$$

As an exercise, show that if, for example,

$$f = v_1 a_1 + v_2 a_2,$$

where v_1 y v_2 are known numbers, then

$$\langle \delta f^2 \rangle = v_1^2 \langle \delta a_1^2 \rangle + v_2^2 \langle \delta a_2^2 \rangle + 2v_1 v_2 \langle \delta a_1 \delta a_2 \rangle.$$

On the other hand, if

$$f = a_1^{v_1} + a_2^{v_2},$$

then

$$\frac{\langle \delta f^2 \rangle}{f^2} = v_1^2 \frac{\langle \delta a_1^2 \rangle}{a_1^2} + v_2^2 \frac{\langle \delta a_2^2 \rangle}{a_2^2} + 2v_1 v_2 \frac{\langle \delta a_1 \delta a_2 \rangle}{a_1 a_2}.$$

A last problem is the following one: we want to make the fit

$$z_i = a_1 + a_2 \cos \omega t_i + a_3 \sin \omega t_i + r_i.$$

If it is assumed that the observation errors at different times are not correlated and that they have de same mean square value, $\langle \delta z_i \delta z_j \rangle = \sigma^2 \delta_{ij}$: (1) show that the best fit is the one that minimizes $r^T r$ (i.e., with equal weights); (2) what condition should the sampling times $\{t_i\}$ fulfill so that

(3) mostrar que si se cumplen esas condiciones y se pasa a coordenadas polares,

(3) show that if those conditions are satisfied and a transformation to polar coordinates,

$$a_2 + ia_3 = me^{i\phi},$$

entonces se cumple que

is made, then it is

$$\langle \delta m^2 \rangle = \langle \delta a_2^2 \rangle, \langle \delta \phi^2 \rangle = \frac{\langle \delta m^2 \rangle}{m^2}, \langle \delta m \delta \phi \rangle = 0.$$

Ajuste de varias series equivalentes

Fitting of several equivalent series

En la práctica, muchas veces se necesita ajustar $l (\geq 1)$ series de n datos Z , mediante una expresión lineal con $m (\leq n)$ parámetros libres A , es decir

In practice, many time it is needed to fit $l (\geq 1)$ series of n data Z , by means of a linear expression with $m (\leq n)$ free parameters A , that is

$$\underbrace{Z}_{n \times l} = \underbrace{F}_{n \times m} \underbrace{A}_{m \times l} + \underbrace{R}_{n \times l},$$

donde los co-errores de las diferentes series de datos se suponen proporcionales

where the co-errors of the different series of data are supposed to be proportional

$$\langle \delta Z_{is} \delta Z_{js} \rangle = \sigma_s^2 C_{ij}.$$

Es importante destacar que las diferentes series, los Z_{is} para diferentes s , **no** son realizaciones del mismo proceso z_i , sino grupos independientes de datos (con igual dependencia en la variable i). Por ejemplo, las estadísticas de precipitación pluvial media, para diferentes meses i en *diferentes* ciudades s . La única razón para considerarlas juntas es de orden práctico, para no tener que invertir varias veces la misma matriz.

It is important to point out that the different series, Z_{is} for different s , are **not** realizations of the same process z_i but independent groups of data (with the same dependence in the variable i). For example, the statistics of mean rainfall, for different months i at *different* cities s . The only reason to consider them together is of practical order, so as not to invert several times the same matrix.

Los coeficientes A se determinan como antes: definiendo las matrices

The coefficient A are determined as before: defining the matrices

$$\begin{aligned} W &= C^{-1} \\ M &= (F^T W F)^{-1} \\ S &= M F^T W, \end{aligned}$$

los parámetros y sus co-errores son

the parameters and their co-errors are

$$\begin{aligned} A &= S Z, \\ \langle \delta A_{\mu s} \delta A_{\nu s} \rangle &= \sigma_s^2 M_{\mu\nu}. \end{aligned}$$

Si no se conocen σ_s^2 se pueden estimar de la relación (23), para cada serie de datos, es decir

In the σ_s^2 are not known, they can be estimated from the relationship (23), for each series of data, that is

$$\{\sigma_s^2\} = \frac{\text{diag}(R^T W R)}{n - m}.$$

La siguiente función de MatLab hace estos cálculos:

The following MatLab function makes these calculations:

```
function [A,R2,M] = CuadMin (Z,F,W);
% Ajuste por cuadrados mínimos: Z(n,:) ~ F(n,m)*A(m,:)
% Uso mínimo: A = CuadMin (Z,F);

% Z: datos; F: funciones de ajuste; A: parámetros;
% W: métrica para residuos [opcional];
% R2: cuadrados (de los residuos) entre el
% número de grados de libertad;
% M: inversa de la métrica para los parámetros
%
% Si no se da W:
%   Co-errores de los A(:,1): M*R2(1),
%   de los A(:,2): M*R2(2), etc.
%
% Si se da W: los R2 deberían ser cercanos a 1
%
[n m]=size(F);
if n<m, F=F'; [n m]=size(F); Z=Z'; end
if size(Z,1)~=n, Z=Z'; end
if nargin<3
M=inv(F'*F); A=M*F'*Z;
if n>m, R2=diag((Z-F*A)'*(Z-F*A))/(n-m); end
else
M=inv (F'*W*F); A=M*F'*W*Z;
if n>m, R2=diag((Z-F*A)'*W*(Z-F*A))/(n-m); end
end
```

Consideraciones y ejemplos finales

Para finalizar, vamos a discutir la importancia de la elección de funciones de ajuste, así como a ejemplificar algunas situaciones típicas que se encuentran en la aplicación de este método.

Ortogonalidad de las funciones de ajuste

La matriz W , la cual define la métrica en el espacio de z y r , puede ser transformada a la matriz unitaria I_n (mediante una diagonalización, seguida de un cambio de las escalas de las componentes z_i). Sin embargo, esta transformación puede ser poco recomendable desde el punto de vista numérico y, además, puede oscurecer el significado de las componentes de z . La única razón para mencionar esta transformación fue señalar que la descomposición de z en los “espacios ortogonales” del ajuste Fa y los residuos r , en el sentido de la ecuación (18), también se puede visualizar como perpendicularidad geométrica, como se muestra en las figuras 1 y 2. Vamos ahora a considerar otro tipo de ortogonalidad que interviene también en (18): la de la contribución de los parámetros a , que corresponde a la métrica (F^TWF) . La presencia o no de esta ortogonalidad sí tiene consecuencias prácticas, como veremos enseguida.

Decimos que las funciones de ajuste, las columnas de F , son ortogonales cuando (F^TWF) es diagonal. Por otra parte, si la matriz F tiene dos columnas iguales (o una columna es igual a una combinación lineal de otras) entonces (F^TWF) es singular, su inversa no existe y no se puede calcular a en forma unívoca

Final considerations and examples

To conclude, we will discuss the importance of the election of fitting functions, as well as to exemplify some typical situations that are encountered in the application of this method.

Orthogonality of the fitting functions

The matrix W , which defines the metric in the space of z and r , can be transformed into the unitary matrix I_n (by means of a diagonalization, followed by a change of scale of the components z_i). However, this transformation might not be a good idea from the numerical point of view and, furthermore, it might obscure the meaning of the components of z . The only reason to mention this transformation was to point out that the decomposition of z in the “orthogonal spaces” of fit Fa and residuals r , in the sense of equation (18), can also be visualized as geometric perpendicularity, like it is shown in figures 1 and 2. We are now going to consider another type orthogonality that is also involved in (18): that one in the contribution of the parameters a , which corresponds to the metric (F^TWF) . The presence or not of this orthogonality does have practical consequences, as we will see next.

We say that the fitting functions, the columns of F , are orthogonal when (F^TWF) is diagonal. If, on the other hand, in the matrix F two columns are equal (or one column is equal to a linear combination of the others) then (F^TWF) is singular, its inverse does not exist and one can neither calculate a in an uni-

ni se puede estimar $\langle \delta a \delta a^T \rangle$. Si dos columnas no son estrictamente iguales, pero sí son muy parecidas entonces $(F^T W F)$ es “casi singular”, aunque sí se puede encontrar su inversa, los coeficientes correspondientes y sus incertidumbres van a tomar valores absurdamente grandes.

En el caso óptimo, ecuación (21), ambas métricas son iguales (o proporcionales) a las inversas de las matrices de co-errores de datos y parámetros, respectivamente, es decir $W = C_z^{-1}$ y $(F^T W F) = C_a^{-1}$. Por lo tanto, si una métrica es diagonal los errores correspondientes no están correlacionados y viceversa. La parte (a) de la figura 3 muestra una situación ideal, donde C_a es proporcional a I_m , para el caso $m = 2$: los errores de los a_μ son de igual magnitud y no están correlacionados.

$$\langle \delta a \delta a^T \rangle = \epsilon_a^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (24)$$

En el caso (b) de la misma figura se muestra el mismo ajuste pero resuelto con las funciones de ajuste que no son ortogonales (en este ejemplo, ambas direcciones forman un ángulo $\pi/2 + \alpha$, con $\alpha \neq 0$): En primer lugar, los nuevos parámetros de ajuste b_μ , son mayores que los a_μ , tanto mayores cuanto más paralelas sean las funciones de ajuste (α cercana a $\pm\pi/2$, en el ejemplo de la figura). Más precisamente, la relación entre ambos conjuntos de parámetros es

$$b_1 = a_1 + a_2 \tan \alpha, \quad b_2 = a_2 \sec \alpha.$$

vocal way, nor can estimate $\langle \delta a \delta a^T \rangle$. If two columns are not strictly equal, but they are very similar then $(F^T W F)$ is “almost singular”, although its inverse can be found, the corresponding coefficients and their uncertainties will take absurdly large values.

In the optimum case, equation (21), both metrics are equal (or proportional) to the inverse of the data and parameters co-error matrices, respectively, namely $W = C_z^{-1}$ and $(F^T W F) = C_a^{-1}$. Therefore, if a metric is diagonal the corresponding errors are not correlated and vice versa. Figure 3a shows an ideal situation, where C_a is proportional to I_m , for the case $m = 2$: the errors of a_μ are of the same magnitude and are not correlated.

Case (b) of the same figure shows the same fit but resolved with fitting functions that are not orthogonal (in this example, both directions form an angle $\pi/2 + \alpha$, with $\alpha \neq 0$): First of all, the new fitting parameters b_μ , are larger than the a_μ , the larger the more parallel the fitting functions are (α near to $\pm\pi/2$, in the example of the figure). More precisely, the relationship between both sets of parameters is

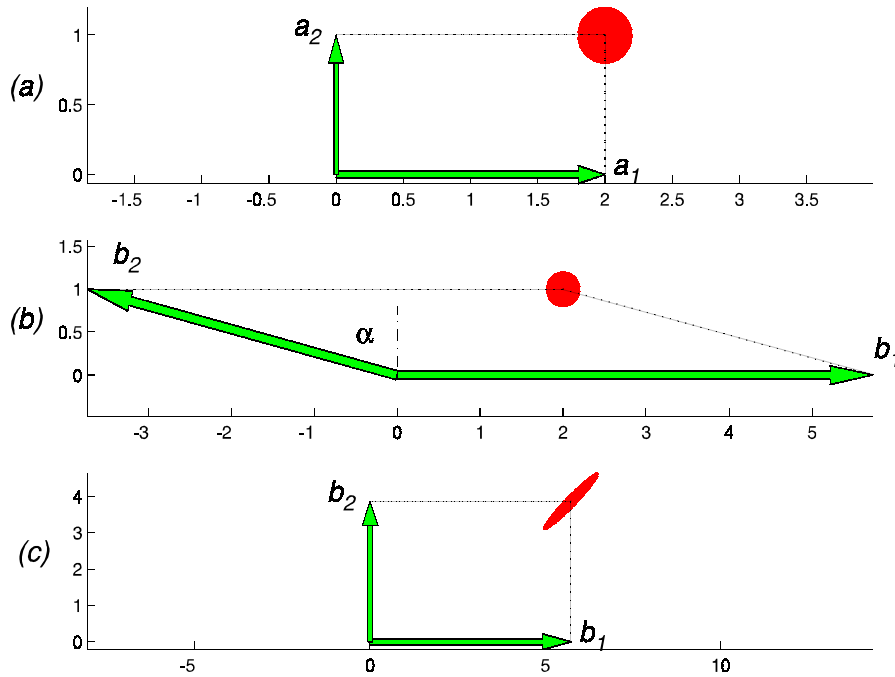


Figura 3. (a) Cuando las funciones de ajuste son ortogonales las incertidumbres de los parámetros (a_1, a_2) no están correlacionadas. **(b)** Con funciones no ortogonales, los parámetros correspondientes (b_1, b_2) son en general mayores (nótese el cambio de escala) y **(c)** sus incertidumbres están fuertemente correlacionadas.

Figure 3. (a) When the adjustment functions are orthogonal the uncertainties of the parameters (a_1, a_2) are not correlated. **(b)** With non orthogonal functions, the corresponding parameters (b_1, b_2) are in general larger (note the change in scale) and **(c)** their uncertainties are strongly correlated.

En segundo lugar, los errores de los b_μ son también mayores que los de los a_μ y están correlacionados; concretamente

$$\langle \delta b \delta b^T \rangle = \frac{\epsilon_a^2}{\cos^2 \alpha} \begin{pmatrix} 1 & \text{sen } \alpha \\ \text{sen } \alpha & 1 \end{pmatrix}; \quad (25)$$

la correlación entre δb_1 y δb_2 es igual a $\text{sen } \alpha$. Esto se aprecia mejor en la parte (c) de la figura 3, donde los ejes son parámetros (b_1, b_2) : la nube de incertidumbre es mucho mayor que en el caso de la figura 3a –con ejes (a_1, a_2) – y su elongación muestra la gran correlación entre δb_1 y δb_2 .

Matemáticamente se ve la clara ventaja de los parámetros (a_1, a_2) frente a los (b_1, b_2) . Sin embargo, los segundos podrían ser los “naturales” para resolver un cierto problema y, por lo tanto, usar los primeros sería un “artificio matemático” que oscurece el significado del ajuste. En este caso (en el que la transformación $a \rightarrow b$ no tiene mucho sentido y $|\alpha|$ es cercano a $\pi/2$) nos estaríamos “auto-engañando” al tomar dos funciones de ajuste tan parecidas. En las siguientes subsecciones vamos a presentar dos ejemplos comunes (ajuste a una recta y a dos componentes de Fourier) en los que el cambio de b a a puede que tenga o no sentido.

Una forma de cuantificar el grado de singularidad de (F¹WF) es mediante el llamado “número de condición” de esa matriz, que es el cociente entre el mayor y el menor de los módulos de sus autovalores. Cuanto mayor es este número, más “singular” es (F¹WF); para el ejemplo de (25), el número de condición es igual a

$$\frac{1 + |\text{sen } \alpha|}{1 - |\text{sen } \alpha|}.$$

(Un lector ignorante de la jerga matemática que lea estas páginas por equivocación se podrá extrañar de que ¡nos interesa propagar los errores y tener la menor condición posible!).

Ajuste a una recta

Al principio de este trabajo comenzamos con el ejemplo más sencillo: n observaciones independientes de la misma variable. Probablemente el segundo ejemplo más usado (a veces, abusado) es el del ajuste de una recta a una serie de n datos $\{z_i\}$ correspondientes a n valores de una variable independiente $\{x_i\}$, es decir

$$z_i = \tilde{a}_1 + \tilde{a}_2 x_i + r_i.$$

con $\sum_i^n = 1/r_i^2$: min. Cambiando el origen y/o la escala de la abscisa, una línea recta sigue siendo una línea recta. Vamos a aprovechar esta propiedad para ilustrar los casos “bien” y “mal” planteados de las figuras 3a, b y c, respectivamente.

In the second place, the errors of the b_μ are also larger than those of the a_μ and they are correlated; specifically

the correlation between δb_1 and δb_2 is equal to $\text{sen } \alpha$. This is appreciated better in part (c) of figure 3c, where the axes are the parameters (b_1, b_2) : the uncertainty cloud is much larger than in the case of figure 3a –with axes (a_1, a_2) – and its elongation shows the high correlation between δb_1 and δb_2 .

Mathematically the advantage of parameters (a_1, a_2) with respect to (b_1, b_2) is clear. However, the latter could be the “natural” ones to solve a certain problem and, therefore, to use the first ones would be a “mathematical artifact” that obscures the meaning of the fit. In this case (in which the transformation $a \rightarrow b$ has no much sense and $|\alpha|$ is close to $\pi/2$) we would be “fooling ourselves” by taking two fitting functions so similar. In the following subsections we will present two common examples (the fit to a straight line and to two Fourier components) in which the change from b to a might or not have sense.

A way to quantify the degree of singularity of (F¹WF) is by means of the so-called “condition number” of that matrix, which is the quotient between the largest and smallest of the modules of their eigenvalues. The larger this number is, the more “singular” is (F¹WF); for the example of (25), the condition number is equal to

(A reader ignorant of the mathematical jargon that reads these pages by mistake might be amazed in knowing that we are interested in propagating error and in having the smallest possible condition!).

Fit to a straight line

At the beginning of this paper we started with the simplest example: n independent observations of the same variable. Probably the second example more oftenly used (sometimes, abused) is that of fitting a straight line to a series of n data $\{z_i\}$ corresponding to n values of an independent variable $\{x_i\}$, namely

with $\sum_i^n = 1/r_i^2$: min. Changing the origin and/or the abscissa scale, a straight line keeps being a straight line. We will take advantage of this property to illustrate the cases “well” and “badly” posed of figures 3a, b and c, respectively.

En primer lugar hacemos un cambio de variable independiente

In the first place we make a change of independent variable:

$$z_i = a_1 + a_2 \eta_i + r_i, \quad x_i = \bar{x} + \Delta_x \eta_i, \quad (26)$$

con $\bar{x} = \sum_{i=1}^n x_i/n$ (i.e., la barra indica media aritmética) y $\Delta_x^2 = \overline{(x_i - \bar{x})^2}$, de manera que $\eta = 0$ y $\Delta_\eta = 1$. Aplicando las fórmulas anteriores, con $W = I_2$ y $F = [1 \ \eta]$, resulta $F^T F = nI_2$ y por lo tanto la solución es

with $\bar{x} = \sum_{i=1}^n x_i/n$ (i.e., the overbar means arithmetic mean) and $\Delta_x^2 = \overline{(x_i - \bar{x})^2}$, so that $\eta = 0$ y $\Delta_\eta = 1$. Applying the previous formulare, with $W = I_2$ y $F = [1 \ \eta]$, resulta $F^T F = nI_2$ and therefore the solution is

$$a_1 = \bar{z}, \quad a_2 = \overline{z\eta},$$

y las incertidumbres son justamente las de (24), con

and the uncertainties are exactly those of (24), with

$$\epsilon_a^2 = \frac{\sum_{i=1}^n r_i^2}{n(n-2)}$$

$= \bar{r}^2/(n-2) = \sigma^2/n$. Como un ejemplo de aplicación de estos resultados, consideremos la interpolación y extrapolación de la recta ajustada, es decir, la función

$= \bar{r}^2/(n-2) = \sigma^2/n$. As an example of the application of these results, let us consider the interpolation and extrapolation of the fitted straight line, that is to say, the function

$$\zeta(\eta) = a_1 + a_2 \eta$$

para cualquier η . Su incertidumbre es $\langle \delta\zeta^2 \rangle = \langle \delta a_1^2 \rangle + \eta^2 \langle \delta a_2^2 \rangle + 2\eta \langle \delta a_1 \delta a_2 \rangle$, o sea

for any η . Its uncertainty is $\langle \delta\zeta^2 \rangle = \langle \delta a_1^2 \rangle + \eta^2 \langle \delta a_2^2 \rangle + 2\eta \langle \delta a_1 \delta a_2 \rangle$, namely

$$\langle \delta\zeta^2 \rangle = \epsilon_a^2 (1 + \eta^2).$$

Las incertidumbres δa_1 y δa_2 no están correlacionadas gracias a haber cambiado de los $\{x_i\}$ a los $\{\eta_i\}$. Los parámetros originales sí están correlacionados, de hecho

The uncertainties δa_1 y δa_2 , are not correlated thanks to having changed from $\{x_i\}$ to $\{\eta_i\}$. The original parameters are correlated, though, in fact it is

$$\begin{aligned} \tilde{a}_1 &= a_1 - a_2 \bar{x} / \Delta_x, \\ \tilde{a}_2 &= a_2 / \Delta_x, \end{aligned}$$

lo que implica la matriz de co-errores

which implies the co-error matrix

$$\langle \delta \tilde{a} \delta \tilde{a}^T \rangle = \epsilon_a^2 \begin{pmatrix} 1 + \bar{x}^2 / \Delta_x^2 & -\bar{x} / \Delta_x^2 \\ -\bar{x} / \Delta_x^2 & 1 / \Delta_x^2 \end{pmatrix};$$

la correlación entre $\delta \tilde{a}_1$ y $\delta \tilde{a}_2$, igual a $-\bar{x} / \sqrt{\bar{x}^2 + \Delta_x^2}$, puede ser muy cercana a 1 o -1 si $|\bar{x}| \gg \Delta_x$.

the correlation between $\delta \tilde{a}_1$ y $\delta \tilde{a}_2$, equal to $-\bar{x} / \sqrt{\bar{x}^2 + \Delta_x^2}$, can be very close to 1 or -1 si $|\bar{x}| \gg \Delta_x$.

Los parámetros \tilde{a}_1 y \tilde{a}_2 no son los de la figura 3b, c (de hecho, podrían tener diferentes unidades). En segundo lugar hacemos entonces el cambio de variable:

The parameters \tilde{a}_1 and \tilde{a}_2 are not those of figure 3b, c (in fact, they could have had different units). In the second place we then make the change of variable:

$$z_i = b_1 + b_2 v_i + r_i, \quad x_i = \sqrt{\bar{x}^2 + \Delta_x^2} v_i, \quad (27)$$

donde obviamente $\bar{x}^2 = \bar{x}^2 + \Delta_x^2$. Claramente es $b_1 = \tilde{a}_1$ y $b_2 = \tilde{a}_2 \sqrt{\bar{x}^2 + \Delta_x^2}$; se puede ver que los nuevos parámetros

where obviously $\bar{x}^2 = \bar{x}^2 + \Delta_x^2$. Clearly it is $b_1 = \tilde{a}_1$ y $b_2 = \tilde{a}_2 \sqrt{\bar{x}^2 + \Delta_x^2}$; it can be seen that new parameters are in

son precisamente los de la figura 3b, c; en particular, sus incertidumbres están dadas por (25) con

$$\tan \alpha = -\frac{\bar{x}}{\Delta_x}$$

(que da el cambio entre las variables independientes como $\eta = v \sec \alpha - \tan \alpha$). Una correlación potencialmente tan grande es, en cierto sentido, artificial ya que puede ser eliminada por un cambio de variables. El siguiente ejemplo muestra un caso donde la correlación entre los parámetros pueden ser muy grande y no se puede eliminar sin perder el sentido del ajuste. De hecho, señala una limitación del método de cuadrados mínimos, que no siempre es bien apreciada.

Análisis de la marea semidiurna

Supongamos que se dispone de los datos de elevación de la superficie del mar $\{z_i\}$ en un puerto, registrados en los tiempos

$$\{t_i\} = \{0, \Delta t, 2\Delta t, \dots, T\},$$

por un instrumento que “filtra” el oleaje y otras oscilaciones de frecuencia alta. Se desea aplicar el método de cuadrados mínimos en la forma

$$z_i = a_1 + a_2 \cos \omega_L t_i + a_3 \sin \omega_L t_i + a_4 \cos \omega_S t_i + a_5 \sin \omega_S t_i + r_i,$$

con $\sum_{i=1}^n r_i^2 : \min$, donde ω_L y ω_S son las frecuencias de las dos componentes principales de la marea semidiurna, lunar y solar respectivamente. Éstas están ligadas por la relación

$$\omega_L = \omega_S - 2\omega_M,$$

donde ω_S y ω_M son las frecuencias aparentes de la rotación solar y del mes lunar, correspondientes a los períodos

$$\begin{aligned} 2\pi/\omega_S &= 12 \text{ h}, \\ 2\pi/\omega_M &= 29\text{d} + 12 \text{ h} + 44 \text{ m} + 2.8 \text{ s}. \end{aligned}$$

El batido entre ω_L y ω_S tiene un periodo π/ω_M de aproximadamente una quincena: en una semana se pasa de “mareas vivas” (suma constructiva) a “mareas muertas” (suma destructiva) y viceversa en la semana siguiente.

Si la duración T del registro es mucho menor que el batido π/ω_M entre ambas componentes, entonces éstas no logran desfasearse lo suficiente como para que sus contribuciones al ajuste estén bien diferenciadas. Con $T \ll \pi/\omega_M$, el ajuste es “mejor” con las dos componentes semidiurnas que con una sola, en el sentido de que $\sum_{i=1}^n r_i^2$ es ligeramente menor en el primer caso (ya que hay más funciones disponibles para minimizarlo). Sin embargo, aunque el ajuste sea un poco mejor con más funciones, estamos engañándonos a nosotros mismos, ya que se está tomando prácticamente la misma función dos veces para

fact those of figure 3b, c; in particular, their uncertainties are given by (25) with

(wich gives the relationship among independent variables as $\eta = v \sec \alpha - \tan \alpha$). A correlation potentially so large is, in a certain sense, artificial since it can be eliminated by a change of variables. The following example shows a case where the correlation among the parameters can be very large and it cannot be eliminated without losing the meaning of the fitting. In fact, it points out a limitation of the least squares method which is not always well appreciated.

Semidiurnal tide analysis

Let us assume that we have observations of sea surface elevation $\{z_i\}$ in port, taken at the times

with an instrument that “filters” the surf and other high frequency oscillations. We want to apply the least squares method in the form

with $\sum_{i=1}^n r_i^2 : \min$, where ω_L and ω_S are the frequencies of the two main components of the semidiurnal tide, of lunar and solar origin, respectively. These are tied by the relationship

where ω_S and ω_M are the apparent frequencies of the solar rotation and of the lunar month, corresponding to the periods

The beat between ω_L and ω_S has a period π/ω_M of approximately a fortnight: in one week the sea goes from “high tide” (constructive interference) to “low tide” (destructive interference) and vice versa in the following week.

If the duration T of the records is much smaller that the beat π/ω_M between both components, then they have no time to get sufficiently out of phase so that their contributions to the fitting are well differentiated. With $T \ll \pi/\omega_M$, the fitting is “better” with two semidiurnal components than with only one, in the sense that $\sum_{i=1}^n r_i^2$ is slightly smaller in the first case (since there are more functions available to minimize it). However, although the fitting is somewhat better with more functions, we are really fooling ourselves, since we are taking practically

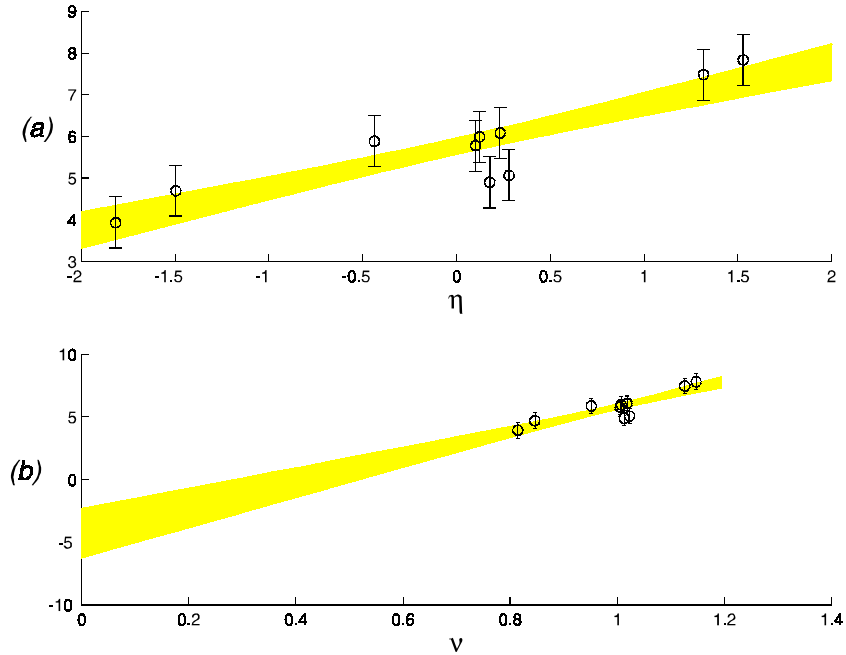


Figura 4. El sombreado muestra la predicción hecha en base el ajuste de los puntos a una línea recta más su incertidumbre estadística. **(a)** Las abscisae de los datos tienen media nula y varianza unitaria; ecuación (26). Ambos parámetros de la recta tienen incertidumbres idénticas y no están correlacionadas (como en el ejemplo de la figura 3a). **(b)** Las abscisae de los datos tienen varianza unitaria pero media diferente de cero; ecuación (27). Los parámetros tienen valores muy grandes e incertidumbres grandes y correlacionadas (como en el ejemplo de la figura 3b,c).

Figure 4. The shading shows the prediction made based on fitting the points to a straight line plus its statistical uncertainty. **(a)** The data abscissae have zero mean and unit variance; equation (26). Both parameters of the straight line have identical uncertainties and are not correlated (like in the example of figure 3a). **(b)** The data abscissae have unit variance but a non-zero mean; equation (27). The parameters have very large values and large and correlated uncertainties (like in the example of figure 3b,c).

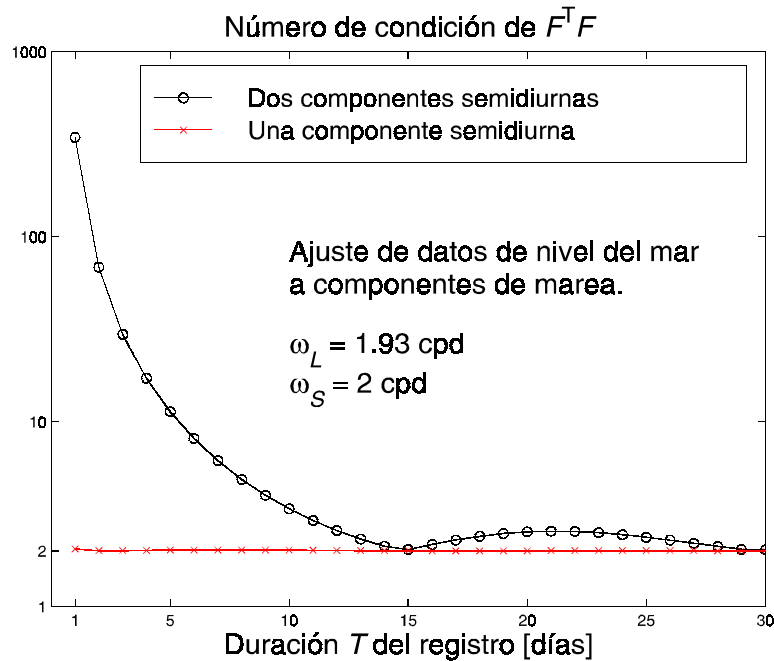


Figura 5. Para ajustar los datos de un registro a dos componentes armónicas es necesario invertir la matriz $(F^T F)$; los co-errores de los parámetros son proporcionales a esa inversa. Cuanto más grande es el número de condición, más cerca está la matriz $(F^T F)$ de ser singular, es decir, de que no exista tal inversa (la situación de las figuras 3b,c).

Figure 5. In order to fit the data from a record to two harmonic components it is necessary to invert the matrix $(F^T F)$; the co-errors of the parameters are proportional to that inverse. The larger the condition number is, the closer the matrix $(F^T F)$ is of being singular, i.e., of not having an inverse (the situation of the figures 3b,c).

optimizarlo: una combinación muy particular de dos sinusoides con amplitudes enormes y un desfase al origen muy preciso logra unos residuos ligeramente inferiores a los del caso con una sola componente. Por supuesto que al extrapolar este ajuste más allá del tiempo T se obtienen resultados absurdos: ¡una semana más tarde se inunda el puerto! Esto corresponde al ejemplo de la figura 3b, c.

En la figura 4 se muestra el número de condición de $(F^T F)$ para $T = 1d, 2d, \dots$ y así hasta un mes, calculado con $\Delta t = 1$ h. Para $T \ll \pi/\omega_M$ el número de condición es muy grande, los resultados serán absurdos. (En la notación de la figura 3, no tiene sentido cambiar a coordenadas a , en lugar de las b , ya que son éstas, ambas componentes semidiurnas de la marea, las que nos interesan). Como comparación, en la figura 4 se muestra también el número de condición de $(F^T F)$ utilizando una sola componente semidiurna, que es prácticamente igual a 2 para todos los T a partir de un día. El número de condición para dos componentes es mínimo (aproximadamente igual a 2) para $T = 15$ d y 30 d, es decir, en uno o dos ciclos (casi) exactos del batido entre ω_L y ω_S , lo que hace que las columnas de F no sólo sean independientes sino también ortogonales: $(F^T F)$ es una matriz diagonal; esto corresponde al caso de la figura 3a.

Variación estacional

Este último caso tiene que ver con estimar la variación estacional de, por ejemplo, el nivel del mar en una localidad. Es común el partir de las medias mensuales de esa variable, como función del mes y año correspondientes. Al tomar esas medias, se filtran oscilaciones de periodo corto –inferiores al mes– y quedan las variaciones estacional (la cual es, por definición, la misma paracada año) e interanual (lo que varía de un año a otro). El paso siguiente consiste en promediar esas medias, sobre todos los años y para cada mes. El resultado son n ($= 12$) valores $\{z_i\}$ correspondientes a los tiempos

$$\{t_i\} = \left\{ \frac{1}{24}, \frac{3}{24}, \frac{5}{24}, \frac{7}{24}, \frac{9}{24}, \frac{11}{24}, \frac{13}{24}, \frac{15}{24}, \frac{17}{24}, \frac{19}{24}, \frac{21}{24}, \frac{23}{24} \right\} T$$

(donde T es igual a un año) y la desviación estándar σ_0 : de todos los datos originales respecto de los z_i . Nuestra hipótesis de partida es que los 12 z_i son una buena representación de la variación estacional; de hecho la vamos a refinar haciendo un ajuste a sinusoides anual y semianual:

$$z_i = a_1 + a_2 \cos \omega t_i + a_3 \sin \omega t_i + a_4 \cos 2\omega t_i + a_5 \sin 2\omega t_i + r_i, \quad (28)$$

con $\omega = 2\pi/T$ y $\sum r_i^2$: min (pesos unitarios). Por otra parte, toda la variación interanual está representada por un sólo número, σ_0 . El ajuste (28) no es, por supuesto, perfecto; los residuos están caracterizados por otro número σ : Para la matriz de errores de los datos se supone la estructura

$$C_z = \sigma^2 I_n + \sigma_0^2 1_n, \quad (29)$$

twice the same function to optimize it: a very peculiar combination of two sinusoids with enormous amplitudes and a very precise relative phase achieves residuals slightly smaller than those of the case with a single component. Of course, by extrapolating this fit beyond time T absurd results are obtained: one week later the port is flooded! This corresponds to the example on figure 3b, c.

The condition number of $(F^T F)$ is shown in figure 4 for $T = 1d, 2d, \dots$ and so on, up to one month, calculated with $\Delta t = 1$ h. For $T \ll \pi/\omega_M$ the condition number is huge, so the results will be absurd. (In the notation of figure 3, to change to coordinates a , instead of b doesn't make sense, since these are the ones of interest: both semidiurnal tide components). As a comparison, in figure 4 the condition number of $(F^T F)$ is also shown using a single semidiurnal component, which is practically equal to 2 for all T starting from one day. The condition number for two components is minimum (approximately equal to 2) for $T = 15$ d and 30 d, namely, (almost) exactly one or two cycles of the beat between ω_L y ω_S , which makes the columns of F not only independent but also orthogonal: $(F^T F)$ is a diagonal matrix; this corresponds to the case in figure 3a.

Seasonal variation

This last case has to do with estimating the seasonal variation of, say, the sea level at a certain place. It is common to start from the monthly means of that variable, as a function of the corresponding month and year. Short period oscillations –less than one month– are filtered by taking these means, remaining but the seasonal variation (the one which is, by definition, the same every year) and the interannual variations (what changes from one year to other one). The following step consists on averaging these means, over all years and for every month. The result are n ($= 12$) values $\{z_i\}$ wich correspond to the times

(where T is equal to one year) and the standard deviation σ_0 : of all the original data relative to the z_i . Our first hypothesis is that the 12 z_i are a good representation of the seasonal variation; in fact we will refine it making a fit to annual and semiannual sinusoids:

with $\omega = 2\pi/T$ and $\sum r_i^2$: min (i.e., using unit weights). On the other hand, the whole interannual variation is represented by a single number, σ_0 . The fit (28) is not perfect, of course: the residuals are characterized by another number σ : The data co-error matrix is assumed to have a structure

donde es bueno recordar que l_n , tiene unos en la diagonal y ceros fuera de ella mientras que todos los componentes de 1_n son iguales a uno. La primera contribución a los errores no está correlacionada entre los meses diferentes, mientras que la segunda tiene una correlación perfecta, ya que modela la variación interanual, en escalas de tiempo mayores que T .

De acuerdo a (21), la matriz de pesos óptima es

$$W_{\text{óptima}} = C_z^{-1} = \frac{1}{\sigma^2} l_n - \frac{n\sigma_0^2}{\sigma^2(\sigma^2 + n\sigma_0^2)} 1_n. \quad (30)$$

Hay aquí, al parecer, dos problemas: Primero, en general, σ no se conoce *a priori* (como en los ejemplos discutidos hasta el momento) pero suele ser menor que σ_0 . (Por ejemplo, para el nivel del mar en el Golfo de California es $\sigma \approx 1.4$ cm y $\sigma_0 \approx 7$ cm, por lo que $n\sigma_0^2/\sigma^2 \approx 300$). Segundo, σ es muy importante: se verá más adelante que es σ –y no σ_0 – quien controla los errores de a_μ para $\mu > 1$ y si se desprecia a σ^2 frente a $n\sigma_0^2$ en (30), la matriz de pesos W que resulta, $\sigma^{-2}(l_n - 1_n)$, es singular.

Sin embargo, la situación no es tan desesperada como parece. Vamos a demostrar que las estimaciones de los parámetros y de las incertidumbres de todos ellos, salvo la de a_1 , son las mismas –para este caso muy particular– utilizando $W = W_{\text{óptima}}$ o $W = l_n$. Se puede demostrar que con los pesos óptimos se obtienen las fórmulas clásicas del análisis de Fourier

$$a_1 = \bar{z}, a_2 = 2\overline{z \cos \omega t}, \dots a_5 = 2\overline{z \sin 2\omega t},$$

donde la barra indica, como antes, la media aritmética y

where it is good to remember that l_n has ones in the diagonal and zeros off it, while all the components of 1_n are equal to one. The first contribution to the errors is not correlated among different months, whilst the second one has a perfect correlation, since it models the interannual variation, in time scales larger than T .

According to (21), the optimum weights matrix is

There are, apparently, two problems: First, σ is generally not known *a priori* (like in the examples discussed so far) but it is usually smaller than σ_0 . (For example, for the sea level in the Gulf of California it is $\sigma \approx 1.4$ cm and $\sigma_0 \approx 7$ cm, and therefore $n\sigma_0^2/\sigma^2 \approx 300$). Second, σ is very important: it will be shown that it is σ –not σ_0 – the one who controls the errors of a_μ for $\mu > 1$ and if one rejects σ^2 in front of $n\sigma_0^2$ in (30), then the weighting matrix W that results, $\sigma^{-2}(l_n - 1_n)$, is singular.

However, the situation is not as desperate as it seems. We are going to demonstrate that the estimations of the parameters and of all their uncertainties, except for that of a_1 , are the same –for this very peculiar case– using $W = W_{\text{óptima}}$ or $W = l_n$. It can be shown that with the optimum weights, the classic formulas of Fourier analysis are obtained

where the bar indicates the arithmetic mean as before, and

$$\langle \delta a \delta a^T \rangle = \begin{pmatrix} \frac{1}{n}\sigma^2 + \sigma_0^2 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{n}\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{n}\sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{n}\sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{n}\sigma^2 \end{pmatrix},$$

donde $n = 12$. Como se puede ver, σ_0^2 sólo influye en la estimación de la incertidumbre de $\langle \delta a_1^2 \rangle$, a pesar de hacer el cálculo tan dificultoso: el número de condición de C_z o de $W_{\text{óptima}}$ es igual a $1 + n\sigma_0^2/\sigma^2$, ¡del orden de 300 para el Golfo de California!

¿Qué sucede al utilizar $W = \sigma^{-2}l_n$, que no son los pesos óptimos? El resultado puede en un principio parecer sorprendente: con $W = \sigma^{-2}l_n$ se obtienen las mismas estimaciones de a y $\langle \delta a \delta a^T \rangle$ que las que se obtienen con $W = W_{\text{óptima}}$. Además resulta

$$\langle r^T r \rangle = (n - m)\sigma^2, \quad (31)$$

where $n = 12$. It is clear that σ_0^2 only influences the estimate of the uncertainty of $\langle \delta a_1^2 \rangle$, in spite of making the calculation so difficult: the condition number of C_z or of $W_{\text{óptima}}$ is equal to $1 + n\sigma_0^2/\sigma^2$, of the order of 300 for the Gulf of California!

What happens when using $W = \sigma^{-2}l_n$, which are not the optimum weights? The result can in principle seem surprising: with $W = \sigma^{-2}l_n$ the same estimates of a and $\langle \delta a \delta a^T \rangle$ are obtained as those obtained with $W = W_{\text{óptima}}$. Moreover, it is expected

donde $m = 5$, expresión que se puede utilizar para estimar $\sigma^2 = (r^T r)/(n - m)$, con los residuos del ajuste. Esto facilita enormemente el cálculo, pues no es necesario trabajar con la matriz “casi singular” $W_{\text{óptima}}$. ¿Por qué es esto? La diferencia entre $W_{\text{óptima}}$ y $\sigma^{-2}I_n$ es proporcional a un operador de proyección: $(n^{-1}1_n)z$ es un vector columna con todos sus elementos iguales al promedio de los z_i , es decir, a_1 por la primera función de ajuste. Ya que $(n^{-1}1_n)$ proyecta en un subespacio de los ajustes posibles, es $(n^{-1}1_n)r = 0$, eso explica el resultado (31). Por otra parte, las otras funciones de ajuste, los cosenos y senos de ωt y de $2\omega t$, tienen un promedio nulo: $(n^{-1}1_n)F$ da cero para todas las columnas de F , salvo la primera. Esto es suficiente para demostrar que las estimaciones de a y $\langle \delta a \delta a^T \rangle$ son las mismas con ambos pesos (excepto $\langle \delta a_1^2 \rangle$, a la cual obviamente le falta la contribución de σ_0^2).

En suma, ya que la variación interanual se supone que actúa de igual forma en todos los meses y que las funciones de ajuste son ortogonales, entonces aquella sólo afecta la estimación del promedio y su incertidumbre. Por ejemplo, el ajuste al nivel total

$$\zeta(t) = a_1 + a_2 \cos \omega t + a_3 \sin \omega t + a_4 \cos 2\omega t + a_5 \sin 2\omega t$$

(para cualquier t) tiene una incertidumbre al cuadrado $\frac{m}{n}\sigma^2 + \sigma_0^2$, es decir

$$\langle \delta \zeta^2 \rangle = \frac{5}{12}\sigma^2 + \sigma_0^2,$$

mientras que el ajuste a la variación (respecto del promedio temporal)

$$\eta(t) = a_2 \cos \omega t + a_3 \sin \omega t + a_4 \cos 2\omega t + a_5 \sin 2\omega t \tag{32}$$

tiene una incertidumbre al cuadrado igual a $\frac{m-1}{n}\sigma^2$, es decir

$$\langle \delta \eta^2 \rangle = \frac{1}{3}\sigma^2, \tag{33}$$

Este ejemplo teórico es, por supuesto, bastante idealizado. Para ilustrar cómo se aplica esta técnica a un caso real, consideremos las observaciones del nivel del mar hechas en Cabo San Lucas (22°53'N 109°54'W), México, entre 1973 y 1998, almacenadas en el banco de datos del Centro de Investigación Científica y Educación Superior de Ensenada (CICESE). Las figuras 6 y 7 muestran el análisis hecho a estos datos, en la forma aquí descrita. En primer lugar, las medias mensuales (los asteriscos en la figura 6) son descompuestas en la forma

$$\zeta(\text{mes, año}) = \bar{\zeta}(\text{mes}) + \zeta'(\text{mes, año});$$

where $m = 5$, expression that can be used to estimate $\sigma^2 = (r^T r)/(n - m)$, with the residuals of the fitting. This facilitates the calculation very much, because it is not necessary to work with the “almost singular” matrix W_{optimum} . Why is this? The difference between W_{optimum} and $\sigma^{-2}I_n$ is proportional to a projection operator: $(n^{-1}1_n)z$ is a column vector with all its elements equal to the average of the z_i , that is to say, a_1 times the first fitting function. Since $(n^{-1}1_n)$ projects into a subspace of the possible fits, it is $(n^{-1}1_n)r = 0$, which explains the result (31). On the other hand, the other fitting functions, the cosines and sines of ωt and of $2\omega t$, have a vanishing average: $(n^{-1}1_n)F$ gives zero for all the columns of F , except for the first one. This is enough to demonstrate that the estimates of a and $\langle \delta a \delta a^T \rangle$ are the same ones with both weights (except for $\langle \delta a_1^2 \rangle$, for which the contribution σ_0^2 is obviously missing).

In sum, since the interannual variation is supposed to act in the same way in every month and that the fitting functions are orthogonal, then it only affects the estimate of the average and its uncertainty. For example, the fit to the whole sea level

(for any t) has an uncertainty square $\frac{m}{n}\sigma^2 + \sigma_0^2$, namely

but the fit to the variation (from the time average)

has a squared uncertainty equal to $\frac{m-1}{n}\sigma^2$, namely

This theoretical example is, of course, highly idealized. To illustrate how this technique is applied to a real case, let us consider the observations of the sea level in Cabo San Lucas (22°53'N 109°54'W), Mexico, between 1973 and 1998, stored in the *Centro de Investigación Científica y Educación Superior (CICESE) de Ensenada* data bank. Figures 6 and 7 show the analysis made to these data, in the form herein described. First of all, the monthly means (asterisks in figure 6) are decomposed in the form

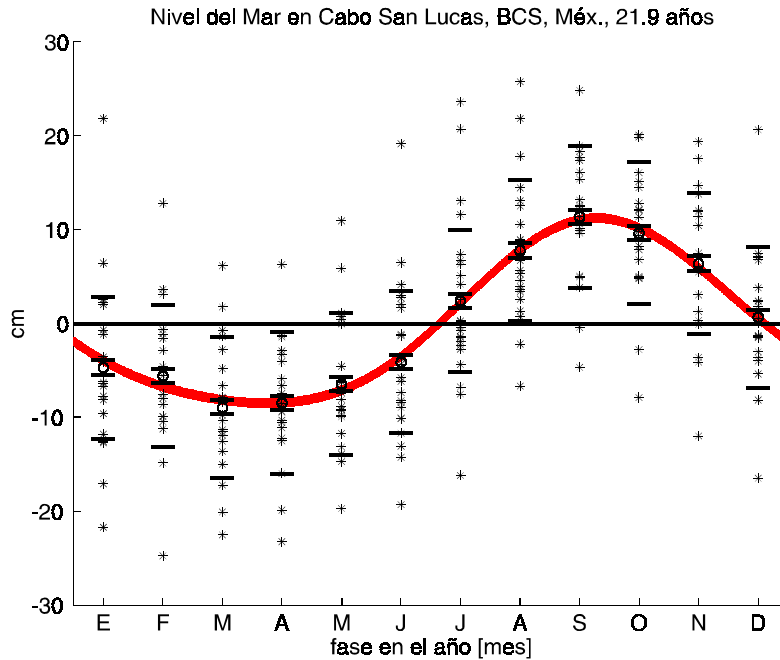


Figura 6. Los asteriscos indican las medias mensuales de nivel del mar y los circulitos muestran los promedios de éstas sobre todos los años disponibles. Cada circulo tiene dos juegos de barras de error: La menor, $\pm\sigma$, indica los errores no correlacionados entre meses diferentes, que determinan la incertidumbre del ajuste estacional (sombreado). La mayor, $\pm\sigma_0$, indica los “errores” perfectamente correlacionados entre meses diferentes, que en realidad tiene que ver con la variación interanual (ver figura 7).

Figure 6. The asterisks indicate the monthly means of sea level and the circles show their average over the available years. Each circle has two sets of error bars: The small one, $\pm\sigma$, indicates the errors uncorrelated between different months, which determine the uncertainty of the seasonal fit (shaded). The large one, $\pm\sigma_0$, indicates the “errors” perfectly correlated between different months, which in fact are related to the interannual variation (see figure 7).

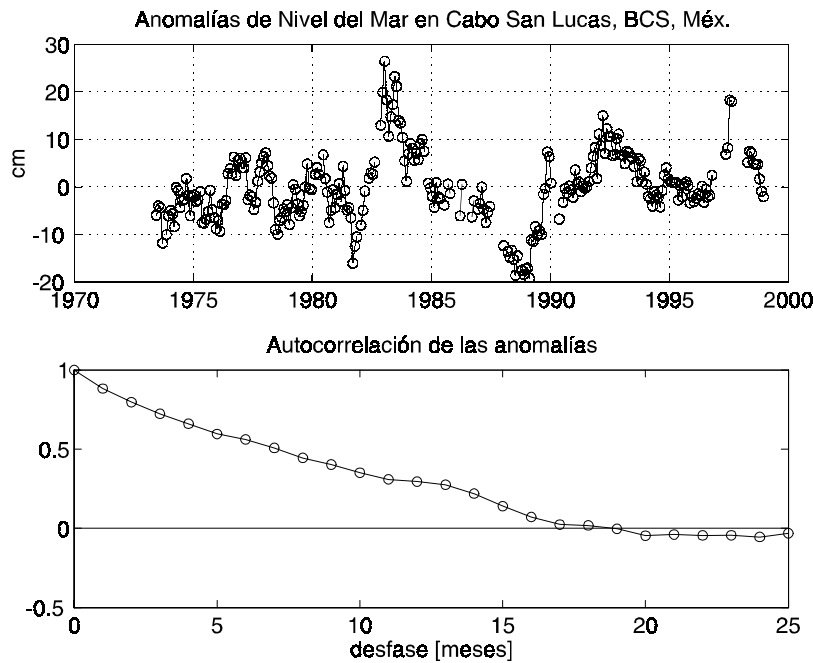


Figura 7. Las anomalías están definidas como la diferencia entre las medias mensuales (para cada mes y año) y sus promedios (para cada mes y sobre todos los años en que hay datos), indicadas por asteriscos y circulitos en la figura 6, respectivamente. Los huecos en la gráfica superior indican datos faltantes.

Figure 7. The anomalies are defined as the difference between the monthly means (for every month and year) and their averages (for every month and over every year with data), indicated by asterisks and circles in the figure 6. The holes in the upper graph indicate missing data.

los doce promedios ζ son indicados por circulitos en la figura 6, mientras que la serie de anomalías ζ' es mostrada en la parte superior de la figura 7. Luego, los ζ son ajustados a la señal estacional (28), bajo la suposición de que la matriz (29) describe adecuadamente la estadística de los errores, donde $\sigma_0 = 7.6$ cm es la desviación estándar de ζ' , mientras que $\sigma = 0.77$ cm es estimada en base a (31). La variación del ajuste (32) y su incertidumbre (33) son mostradas por el sombreado de la figura 6. Las amplitudes de las sinusoides anual y semi-anual son estimadas como $|a_2 + ia_3| = 9.8 \pm 0.3$ cm y $|a_4 + ia_5| = 1.5 \pm 0.3$ cm, respectivamente.

Es bien claro que sin la hipótesis (29) entonces la incertidumbre de la variación del nivel del mar (32) sería un orden de magnitud mayor $\sigma/\sqrt{3}$. Por ejemplo, si se hiciera el ajuste estacional directamente a las medias ζ (mes, año) entonces sería $\sqrt{\langle \delta\eta^2 \rangle} = O(\sigma_0)$, ¡similar a la amplitud anual! ¿Es (29) una hipótesis razonable? La correlación (temporal) entre $\zeta'(t)$ y $\zeta'(t + \tau)$, parte inferior de la figura 7, disminuye sensiblemente para desfases τ del orden de un año, de manera que considerar a ζ' como un ruido con contribución a $\langle \delta\zeta\delta\zeta^T \rangle$ igual a $\sigma_0^2 1_{12}$, sí parece una hipótesis razonable *para los efectos del ajuste estacional* de ζ . Por supuesto que en escalas de tiempo mayores, ζ' muestra una estructura (parte superior de la figura 7) que se puede asociar a procesos climáticos conocidos. (Esto se refleja en forma de oscilaciones de periodo largo en la prolongación de parte inferior de la figura 7, para desfases τ del orden de varios años).

La estadística por sí sola no asegura que el ajuste tenga sentido físico. Como dijo G.K. Chesterton:

“La estadística debería ser usada de la misma manera que un ebrio utiliza un poste de luz ... más como apoyo que por la iluminación”. (El por qué estaba interesado el gran novelista inglés en la estadística, no lo sé).

Al utilizar el ajuste estacional de la figura 6 y las anomalías de la figura 7 en un contexto científico más amplio se justificará o condicionará la validez del análisis estadístico.

Conclusiones

El algoritmo lineal que reduce un conjunto de datos a uno menor, de parámetros de ajuste, minimizando ciertos residuos,

$$z \rightarrow a, \quad \text{min } \|r\|$$

es directo y bien conocido. La misma transformación se aplica a los errores $\delta z \rightarrow \delta a$, las desviaciones respecto de los “valores verdaderos” (obviamente desconocidos); de manera que se puede pasar de estadísticas de los datos (incertidumbre estimadas, etc.) a estadísticas de los parámetros.

Sin embargo, comúnmente este programa no puede ser llevado a cabo porque no se conocen *a priori* los valores cuadráticos medios de los errores de las observaciones ni sus

the twelve averages ζ are indicated by circles in figure 6, whilst the anomalies' series ζ' are shown in the upper part of figure 7. Then, the ζ are fitted to the seasonal signal (28), under the assumption that the matrix (29) describes appropriately the errors statistics, where $\sigma_0 = 7.6$ cm is the standard deviation of ζ' , and $\sigma = 0.77$ cm is estimated from (31). The time variation of the fit (32) and its uncertainty (33) are shown by the shading in figure 6. The amplitudes of the annual and semiannual sinusoids are estimated as $|a_2 + ia_3| = 9.8 \pm 0.3$ cm and $|a_4 + ia_5| = 1.5 \pm 0.3$ cm, respectively.

It is clear that without hypothesis (29) the uncertainty of the variation (32) in the sea level would be an order of magnitude larger than $\sigma/\sqrt{3}$. For example, if the seasonal fit were made directly to the means ζ (month, year), then it would be $\sqrt{\langle \delta\eta^2 \rangle} = O(\sigma_0)$, of the order of the annual amplitude! Is (29) a reasonable hypothesis? The (time) correlation between $\zeta'(t)$ and $\zeta'(t + \tau)$, lower part of figure 7, diminishes sensibly for lags τ of the order of one year, so that to consider ζ' as noise with a contribution to $\langle \delta\zeta\delta\zeta^T \rangle$ equal to $\sigma_0^2 1_{12}$, does seem to be a reasonable hypothesis for *the effects of the seasonal fit of ζ* . Of course that, in longer time scales, ζ' shows a structure (upper part of figure 7) that can be associated to well-known climatic processes. (This is reflected as long-period oscillations in the continuation of the lower part of figure 7, for lags τ of the order of several years).

The statistic by itself cannot assure that the fitting makes physical sense. As G.K. Chesterton said, “Statistics should be used as drunk uses a lamp post ... more for support than for illumination”. (I don't know why the great English novelist was interested in statistics).

The use of the seasonal fit in figure 6 and the anomalies in figure 7 in a wider scientific context will justify or condition the validity of the statistical analysis.

Conclusions

The linear algorithm that reduces a set of data to a smaller one, of fitting parameters, by minimizing certain residuals

is quite direct and very well-known. The same transformation is applied to the errors $\delta z \rightarrow \delta a$, the deviations from their “true values” (obviously unknown); this way one can to pass from data statistics (estimated uncertainties, etc.) to parameters statistics.

However, this program cannot usually be carried out because the mean square of the observation errors, their correlations, etc. are not known *a priori*. In this case, using the

correlaciones. En este caso, se suele utilizar la magnitud de los residuos mínimos para estimar esas estadísticas desconocidas. Sin embargo, para poder hacerlo, es necesario postular ciertas hipótesis; las más comunes son:

1. Los errores de los datos son totalmente aleatorios (no hay errores sistemáticos): el promedio estadístico de estos errores es nulo.
2. Los errores de los datos no están correlacionados.
3. Los valores cuadráticos medios de los errores de los datos son iguales.
4. El valor verdadero de los datos tiene un ajuste perfecto (el promedio estadístico de los residuos es nulo).

Aunque las hipótesis pueden cambiar (v.g., si los datos tienen diferentes unidades, claramente hay que modificar la tercera), no se puede prescindir de ellas y por lo tanto condicionan la validez de las conclusiones. Por ejemplo, si se hacen ajustes con diferente tipo y/o número de funciones, la cuarta hipótesis no puede ser válida para todos y cada uno de los ajustes.

El método de cuadrados mínimos es sencillo y eficaz: dadas las observaciones y las funciones de ajuste calcula los parámetros que minimizan los residuos. Si se agregan más funciones de ajuste, se lograrán –en general– residuos menores, pero ¿realmente se gana información agregando más parámetros? Si las funciones de ajuste no son ortogonales, o sea, si cada una no aporta una posibilidad de variación totalmente novedosa, entonces el ajuste puede dar resultados absurdos, con parámetros demasiado grandes. La buena noticia es que el método no sólo da valores muy grandes de los coeficientes sino que también lo hace para sus incertidumbres; errores grandes y muy correlacionados hacen sospechar algo no muy bien planteado.

Matriz óptima de pesos

En esta sección se señala cómo demostrar la ecuación (21). De la definición (17) de M resulta $M(F^T W F) = I_m$. Calculando su primera variación, para una variación infinitesimal de la matriz de pesos, $\delta W \ll W$, resulta $(\delta M) F^T W F + M F^T (\delta W) F = 0$, o sea

$$\delta M = -M F^T (\delta W) F M.$$

Usando este resultado en la primera variación de $S = M F^T W$ se puede demostrar que

$$\delta S = M F^T (\delta W) Q.$$

Finalmente, reemplazando en la primera variación de $C_a = S C_z S^T$, o sea $\delta C_a = (\delta S) C_z S^T + S C_z (\delta S)^T$, se obtiene

$$\delta C_a = M F^T (\delta W) \boxed{Q C_z W F} M + \text{transpuesta}.$$

magnitude of the (minimum) residuals is common to estimate those unknown statistics. However, in order to be able to make that, certain hypotheses have to be postulated; the most common are:

1. Data errors are entirely random (there are no systematic errors): the statistical average of these errors vanishes.
2. Data errors are not correlated.
3. Mean square data errors are equal.
4. The true value of the data has a perfect fit (the statistical average of the residuals it is null).

Although these hypotheses may change (e.g., if data have different units, clearly the third one has to be modified), one is not able to do without of them and therefore they condition the validity of the conclusions. For example, if fits are made with different type and/or number or functions, the fourth hypothesis cannot be valid for every one of those fits.

The least squares method is simple and effective: given the data and the fitting functions it calculates the parameters that minimize the residuals. If more fitting functions are added, smaller residuals will be achieved –in general– but, Has new information been really gained by adding more parameters? If the fitting functions are not orthogonal, namely, if each one does not contribute a completely novel possibility of variation, then the fit can yield absurd results, with too big parameters. The good news is that the method gives huge values not only for the coefficients but also for their uncertainties; large and very correlated errors make one suspect that something is not very well posed.

Optimum weights matrix

In this section it is shown how to demonstrate equation (21). From the definition (17) of M it is $M(F^T W F) = I_m$. Calculating the first variation, for an infinitesimal variation of the weights matrix, $\delta W \ll W$, it is $(\delta M) F^T W F + M F^T (\delta W) F = 0$, and therefore

Using this result in the first variation of $S = M F^T W$ it can be shown that

Finally, replacing in the first variation of $C_a = S C_z S^T$, namely $\delta C_a = (\delta S) C_z S^T + S C_z (\delta S)^T$, it is obtained

La parte importante se ha puesto en un recuadro: si $C_z W \propto I_n$, los “pesos óptimos” de (21), entonces $\delta C_a = 0$, ya que $QF = 0$ (pues Q proyecta hacia el espacio de los residuos, la parte que no es expandida por las funciones F).

Notas

¹ Esta desviación estándar resulta de suponer que los errores δz_i tienen una distribución gaussiana multivariada. En tal caso $\|r\|^2$ es una variable aleatoria con una distribución χ^2 con $n - 1$ grados de libertad. Probablemente es mejor no caer en tanto detalle y simplemente tomar a (11) como una indicación “típica”.

² Luego z , a y r son vectores columna. Hay que estar habituado con las reglas del producto de matrices y la acción de transponer (indicada por T). Por ejemplo, $z^T z$ y $a^T a$ son números, mientras que $z z^T$ y $a a^T$ son matrices cuadradas de $n \times n$ y $m \times m$, respectivamente. Además, al transponer un producto $(Fa)^T = a^T F^T$.

El lector que no guste de la notación matricial puede reescribir todas las fórmulas con subíndices. La multiplicación de matrices es equivalente a la suma sobre los índices interiores. La traza de una matriz (la que debe ser una cuadrada) es equivalente a la suma sobre los índices exteriores y se indica por $\text{tr}(\dots)$.

³ Ante una variación infinitesimal δa en $r^T W r = (z^T - a^T F^T) W (z - Fa)$ es:

$$\begin{aligned} \delta(r^T W r) &= -\delta a^T F^T W (z - Fa) - (z - Fa)^T W F \delta a \\ &= -2(\delta a)^T F^T W (z - Fa) . \end{aligned}$$

Ya que δa^T es arbitrario, debe ser $F^T W (z - Fa) = 0$, cuya solución es (17).

⁴ Los símbolos 0_n , 1_n e I_n indican una matriz con puros ceros, una con puros unos y la matriz unidad (unos en la diagonal y ceros fuera de ella), todas de $n \times n$.

⁵ De (15) es

$$\begin{aligned} z^T W z &= (Fa)^T W F a + r^T W r \\ &\quad + r^T W F a + (Fa)^T W r . \end{aligned}$$

Una forma de demostrar (18), o sea que los términos cruzados –de la segunda línea– son nulos, es escribiéndolos como

$$r^T W F a = z^T Q^T W P z$$

y su transpuesta, y utilizando $Q^T W P = W Q P = 0_n$.

The important part is framed: if $C_z W \propto I_n$, the “optimum weights” of (21), then $\delta C_a = 0$, since $QF = 0$ (because Q projects into the residuals’ space, part which cannot be expanded by the functions F).

Notas

¹ This standard deviation results from assuming that the errors δz_i have a multivariate Gaussian distribution. In such a case $\|r\|^2$ is a random variable with a χ^2 distribution with $n - 1$ degrees of freedom. Probably, it is better not to fall into so much detail, and simply take (11) as a “typical” indication.

² Then z , a and r are column vectors. One must be familiar with the rules of matrix multiplication and transposing (indicated by T). For instance, $z^T z$ and $a^T a$ are numbers, while $z z^T$ and $a a^T$ are square $n \times n$ and $m \times m$ matrices, respectively. Also, the way to transpose a product is $(Fa)^T = a^T F^T$.

The reader who does not prefer matrix notation can rewrite all formulae with subscripts. Matrix multiplication is equivalent to sum over the internal subscripts. The trace of a matrix (which has to be a squared one) is equivalent to the sum over the external subscript and it is indicated by $\text{tr}(\dots)$.

³ Making an infinitesimal variation δa in $r^T W r = (z^T - a^T F^T) W (z - Fa)$ it is:

Since δa^T is arbitrary, it must be $F^T W (z - Fa) = 0$, whose solution is (17).

⁴ The symbols 0_n , 1_n e I_n indicate a matrix with all zeroes, one with all ones, and the unit matrix (ones in the diagonal and zero off the diagonal), all $n \times n$.

⁵ From (15) it is

One way to demonstrate (18), namely, that the cross terms – in the second line– vanish, is writing them as

and its transposed, and using $Q^T W P = W Q P = 0_n$.

⁶ De (20) es

⁶ From (20) it is

$$\begin{aligned}\langle \delta a \delta a^T \rangle &= \langle MF^T W \delta z \delta z^T W F M \rangle \\ &= MF^T W (\delta z \delta z^T) W F M,\end{aligned}$$

ya que W , M y F no son aleatorias, y las dos primeras son simétricas.

since W , M and F are not random, and the first two are symmetric.

⁷

$$\begin{aligned}\langle \|\delta z\|^2 \rangle &= \langle \delta z^T W \delta z \rangle = \text{tr} \langle \delta z \delta z^T W \rangle = \text{tr}(C_z W) = \text{tr}(I_n) = n \\ \langle \|F \delta a\|^2 \rangle &= \langle \delta z^T P^T W P \delta z \rangle = \langle \delta z^T W P^2 \delta z \rangle = \text{tr} \langle \delta z \delta z^T W P \rangle = \text{tr}(C_z W P) = \text{tr}(P) = m.\end{aligned}$$

Agradecimientos

Este es el tipo de tema que uno no recuerda muy bien dónde y cómo lo fue aprendiendo, ni tampoco está claro qué aportación original se ha hecho, si es que ésta existe. Por lo tanto, quiero agradecer a todos aquellos –los que recuerdo y los demás– que me enseñaron a mirar a los datos con esperanza y confianza. También quiero agradecer a José Ochoa por revisar el manuscrito y a Ignacio González por los datos de nivel del mar en Cabo San Lucas. Este trabajo ha sido financiado por el presupuesto interno del CICESE.

Acknowledgements

This is the kind of topic one neither remembers very well where and how has learn it, nor it is clear what is the original contribution that has been made –if it exists at all. Therefore, I want to thank all those people –those that I remember and the other ones too– that taught me to look at data with hope and distrust. I also want to thank José Ochoa for reading the manuscript and to Ignacio González for the sea level data from Cabo San Lucas. This work has been funded by CICESE.

English translation by the author.
