Optmal Estimation

Kenichi Kanatani
Professor Emeritus, Okayama University, Japan

Synoyms

Optimal parameter estimation

Related Concepts

- ▶ Maximum Likelihood Estimation
- ▶ Ellipse Fitting, Fundamental Matrix
- ▶ Homography
- ▶ KCR Lower Bound
- ▶ Hyper-renormalization

Definition

Optimal estimation in the computer vision context refers to estimating the parameters that describe the underlying problem from noisy observation. The estimation is done according to a given criterion of optimality, for which maximum likelihood is widely accepted. If Gaussian noise is assumed, it reduces to minimizing the Mahalanobis distance. If furthermore the Gaussian noise has a homogeneous and isotropic distribution, the procedure reduces to minimizing what is called the reprojection error.

Background

One of the central tasks of computer vision is the extraction of 2D/3D geometric information from noisy image data. Here, the term *image data* refers to values extracted from images by image processing operations such as edge filters and interest point detectors. Image data are said to be *noisy* in the sense that image processing operations for detecting them entail uncertainty to some extent.

For optimal estimation, a statistical model of observation needs to be introduced. Let $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ be the observed image data. The standard model is to view each datum $\boldsymbol{x}_\alpha$ as perturbed from its true value $\bar{\boldsymbol{x}}_\alpha$ by $\Delta\boldsymbol{x}_\alpha$, which is assumed to be independent Gaussian noise of mean $\boldsymbol{0}$ and covariance matrix $V[\boldsymbol{x}_\alpha]$. Then, maximum likelihood is equivalent to the minimization of the *Mahalanobis distance*

$$I = \sum_{\alpha=1}^{N} \langle \bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha, V[\boldsymbol{x}_\alpha]^{-1}(\bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha)\rangle, \qquad (1)$$

with respect to the true values $\bar{\boldsymbol{x}}_\alpha$ subject to given knowledge about them. Hereafter, $\langle \boldsymbol{a}, \boldsymbol{b}\rangle$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

If the noise is homogeneous and isotropic, in which case $V[\boldsymbol{x}_\alpha] = c\boldsymbol{I}$ for all α for some constant c and the identity matrix $\boldsymbol{I}$, the Mahalanobis distance I is equivalent to the sum of the squares of the geometric distances between the observations $\boldsymbol{x}_\alpha$ and their true values $\bar{\boldsymbol{x}}_\alpha$. In this case, I is often referred to as the *reprojection error*. That name originates from the following intuition: In inferring the 3D structure of the scene from its projected images, maximum likelihood under homogeneous and isotropic Gaussian noise means *reprojecting* the inferred 3D structure onto the images and minimizing the square distance between the *reprojection* of the solution and the projection of the scene. Reprojection error minimization is also referred to as *geometric fitting*.

Theory

The estimation procedure depends on the way the knowledge about true values $\bar{\boldsymbol{x}}_\alpha$ is represented. A typical approach is to introduce some function $\boldsymbol{g}(\boldsymbol{t}, \boldsymbol{\theta})$ to express $\bar{\boldsymbol{x}}_\alpha$ in a parametric form

$$\bar{\boldsymbol{x}}_\alpha = \boldsymbol{g}(\boldsymbol{t}_\alpha, \boldsymbol{\theta}), \qquad (2)$$

where $\boldsymbol{t}_\alpha$ is a control variable that specifies the identity of the αth datum, and $\boldsymbol{\theta}$ is an unknown parameter that specifies the underlying structure. After (2) is substituted, the Mahalanobis distance I becomes a function of $\boldsymbol{\theta}$ alone, which is then minimized with respect to $\boldsymbol{\theta}$. This is the standard approach in the traditional statistic estimation framework and also known as *regression*.

This parametric approach, however, is quite limited in computer vision applications. Often, no such knowledge as (2) is available about the true values $\bar{\boldsymbol{x}}_\alpha$ except that that they satisfy some implicit equations of the form

$$F^{(k)}(\boldsymbol{x}, \boldsymbol{\theta}) = 0, \qquad k = 1, ..., L. \qquad (3)$$

The unknown parameter $\boldsymbol{\theta}$ allows one to infer the 2D/3D shape and motion of the objects observed in the images.

This type of estimation leads to some theoretical problems. Usually, no restriction is imposed on the true values $\bar{\boldsymbol{x}}_\alpha$ except that they should satisfy (3). This is called the *functional model*. One could alternatively introduce some statistical model according to which the true values $\bar{\boldsymbol{x}}_\alpha$ are "sampled." This model is called *structural*. The distinction is crucial

when one considers limiting processes in the following sense. Traditional statistical analysis mainly focuses on the asymptotic behavior as the number of observations increases to ∞. This is based on the reasoning that the mechanism underlying noisy observations would better reveal itself as the number of observations increases (*the law of large numbers*) while the number of available data is limited in practice. So, the estimation accuracy vs. the number of data is a major concern. In this light, efforts have been made to obtain a consistent estimator in the sense that the solution approaches its true value in the limit $N \to \infty$ of the number N of the data.

In computer vision applications, in contrast, one cannot *repeat* observations. One makes an inference given a single set of images, and how many times one applies image processing operations, the result is always the same, because standard image processing algorithms are deterministic and no randomness is involved. This is in a stark contrast to conventional statistical problems where observations are viewed as "samples" from potentially infinitely many possibilities; we could obtain, by repeating observations, different values originating from unknown, uncontrollable, or unmodeled causes, which is called *noise* as a whole.

In vision problems, the accuracy of inference deteriorates as the uncertainty of image processing operations increases. Thus, the inference accuracy vs. the uncertainty of image operations, which is called *noise* for simplicity, is a major concern. Usually, the noise is very small, often subpixel levels. In light of this observation, it has been pointed out that in image domains the *consistency* of estimators should more appropriately be defined by the behavior in the limit $\sigma \to 0$ of the noise level σ [1, 2]. The functional model suits this purpose. If the error behavior in the limit of $N \to \infty$ were to be analyzed, one needs to assume some structural model that specifies how the statistical characteristics of the data depend on N. However, it is difficult to predict the noise characteristics for different N. Image processing filters usually output a list of points or lines or their correspondences along with their confidence values, from which only those with high confidence are used. If a lot of data are to be collected, those with low confidence need to be included, but their statistical properties are hard to estimate, since such data are possibly misdetections. This is the most different aspect of image processing from laboratory experiments, in which any number of data can be collected by repeated trials.

Maximum Likelihood with Implicit Constraints

Maximum likelihood based on the functional model is to minimize the Mahalanobis distance (1) subject to implicit constraints in the form of (3). In statistics, maximum likelihood is criticized for its lack of consistency. In fact, estimation of the true values $\bar{\boldsymbol{x}}_\alpha$, called *nuisance parameters* when viewed as parameters, is not consistent as $N \to \infty$ in the maximum likelihood framework [3]. However, the lack of consistency has no realistic meaning in vision applications as explained above. On the contrary, maximum likelihood has very desirable properties in the limit $\sigma \to 0$ of the noise level σ: the solution is *consistent* in the sense that it converges to the true value as $\sigma \to 0$ and *efficient* in the sense that its covariance matrix approaches a theoretical lower bound as $\sigma \to 0$ [1, 2].

According to the experience of many vision researchers, maximum likelihood is known to produce highly accurate solutions. A major concern is its computational burden, because maximum likelihood usually requires complicated nonlinear optimization. The standard approach is to express each of $\bar{\boldsymbol{x}}_\alpha$ explicitly in terms of $\boldsymbol{\theta}$ by introducing some auxiliary parameters, or nuisance parameters. After all the expressions are substituted back into (1), the Mahalanobis distance I becomes a function of $\boldsymbol{\theta}$ and the nuisance parameters. Then, this joint parameter space, which usually has very high dimensions, is searched for the minimum. This approach is called *bundle adjustment*, a term originally used by photogrammetrists. This is very time consuming, in particular if one seeks a globally optimal solution by searching the entire parameter space exhaustively.

Linear Reparameterization

In many important vision applications, the problem can be reparameterized to make the functions $F^{(k)}(\boldsymbol{x}, \boldsymbol{\theta})$ linear in $\boldsymbol{\theta}$ (but generally nonlinear in $\boldsymbol{x}$), allowing one to write (3) as

$$\langle \boldsymbol{\xi}^{(k)}(\boldsymbol{x}), \boldsymbol{\theta} \rangle = 0, \qquad k = 1, ..., L, \tag{4}$$

where $\boldsymbol{\xi}^{(k)}(\boldsymbol{x})$ represents a nonlinear mapping of $\boldsymbol{x}$. This formalism covers many fundamental problems of computer vision including fitting a parametric curve such as a line, an ellipse, and a polynomial curve to a noisy 2D point sequence or a parametric surface such as a plane, an ellipsoid, and a polynomial surface to a noisy 3D point sets and computing the fundamental matrix or the homography from noisy point correspondences over two images [4, 5]. For this type of problem, a popular alternative to bundle adjustment is minimization of a function of $\boldsymbol{\theta}$ alone, called the *Sampson error*. Let us abbreviate $\boldsymbol{\xi}^{(k)}(\boldsymbol{x}_\alpha)$ to $\boldsymbol{\xi}^{(k)}_\alpha$.

The first order variation of $\boldsymbol{\xi}_\alpha^{(k)}$ by noise is

$$\Delta\boldsymbol{\xi}_\alpha^{(k)} = \boldsymbol{T}_\alpha^{(k)}\Delta\boldsymbol{x}_\alpha, \quad \boldsymbol{T}_\alpha^{(k)} \equiv \left.\frac{\partial\boldsymbol{\xi}^{(k)}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha}. \quad (5)$$

Define the covariance matrices of $\boldsymbol{\xi}_\alpha^{(k)}$, $k = 1, ..., L$, by

$$\begin{aligned} V^{(kl)}[\boldsymbol{\xi}_\alpha] &= E[\Delta\boldsymbol{\xi}_\alpha^{(k)}\Delta\boldsymbol{\xi}_\alpha^{(l)\top}] \\ &= \boldsymbol{T}_\alpha^{(k)}E[\Delta\boldsymbol{x}_\alpha\Delta\boldsymbol{x}_\alpha^\top]\boldsymbol{T}_\alpha^{(l)\top} \\ &= \boldsymbol{T}_\alpha^{(k)}V[\boldsymbol{x}_\alpha]\boldsymbol{T}_\alpha^{(l)\top}, \end{aligned} \quad (6)$$

where $E[\,\cdot\,]$ denotes expectation. The Sampson error that approximates the minimum of the Mahalanobis distance I subject to the constraints in (4) has the form

$$K = \sum_{\alpha=1}^{N}\sum_{k,l,=1}^{L} W_\alpha^{(kl)}\langle\boldsymbol{\xi}_\alpha^{(k)},\boldsymbol{\theta}\rangle\langle\boldsymbol{\xi}_\alpha^{(l)},\boldsymbol{\theta}\rangle, \quad (7)$$

where $W_\alpha^{(kl)}$ is the (kl) element of $(\boldsymbol{V}_\alpha)_r^-$. Here, $\boldsymbol{V}_\alpha$ is the matrix whose (kl) element is

$$\boldsymbol{V}_\alpha = \Big(\langle\boldsymbol{\theta}, V^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}\rangle\Big), \quad (8)$$

where the true data values $\bar{\boldsymbol{x}}_\alpha$ in the definition of $V^{(kl)}[\boldsymbol{\xi}_\alpha]$ are replaced by their observations $\boldsymbol{x}_\alpha$. The operation $(\,\cdot\,)_r^-$ denotes the pseudoinverse of truncated rank r, (i.e., with all eigenvalues except the largest r replaced by 0 in the spectral decomposition), and r is the rank of $\boldsymbol{V}_\alpha$, which is equal to the number of independent equations of (4). The name Sampson error stems from the classical ellipse fitting scheme [6].

The Sampson error (7) can be minimized by various means including the *FNS* (*Fundamental Numerical Scheme*) [7] and the *HEIV* (*Heteroscedastic Errors-in-Variable*) [8]. It can be shown that the exact maximum likelihood solution can be obtained by repeating Sampson error minimization, each time modifying the Sampson error so that in the end the modified Sampson error coincides with the Mahalanobis distance [9, 5]. It turns out that in many practical applications the solution that minimizes the Sampson error coincides with the exact maximum likelihood solution up to several significant digits; usually, two or three rounds of Sampson error modification are sufficient.

It can be shown that the covariance matrix $V[\hat{\boldsymbol{\theta}}]$ of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ satisfies under some general conditions the inequality

$$V[\hat{\boldsymbol{\theta}}] \succ \Big(\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)}\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}\Big)_r^-, \quad (9)$$

where $\bar{\boldsymbol{\xi}}_\alpha^{(k)}$ are the trues value of $\boldsymbol{\xi}_\alpha^{(k)}$, and $\bar{W}_\alpha^{(kl)}$ is the value of $W_\alpha^{(kl)}$ defined earlier evaluated for the true values of $\boldsymbol{\xi}_\alpha^{(k)}$ and $\boldsymbol{\theta}$. The symbol $\succ$ means that the left-hand side minus the right-hand side is positive semidefinite. The right-hand side of (9) is called the *KCR* (*Kanatani-Cramer-Rao*) *lower bound* [1, 2, 5]. It can be shown that the covariance matrix of Sampson error minimization solution coincides with this bound in the leading order in the noise level [1, 2].

Algebraic methods

Recently, there has been a remarkable progress in the study of algebraic methods. By "algebraic methods," we mean we solve some "algebraic equations" (directly or iteratively), rather than minimizing some cost function such as the reprojection error. Originally, algebraic methods were thought of as an auxiliary to maximum likelihood and used for initialization of maximum likelihood iterations. In the last decade, however, it has been found that some algebraic methods outperform maximum likelihood in accuracy [5].

Algebraic methods solve a nonlinear equation in the form

$$\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{N}\boldsymbol{\theta}, \quad (10)$$

with

$$\boldsymbol{M} = \sum_{\alpha=1}^{N}\sum_{k=1}^{L} W_\alpha^{(kl)}\boldsymbol{\xi}_\alpha^{(k)}\boldsymbol{\xi}_\alpha^{(k)\top}, \quad (11)$$

where $W_\alpha^{(kl)}$ are some weights that depend on $\boldsymbol{\theta}$. Various methods with different names arise according to the choice of the weights $W_\alpha^{(kl)}$ and the matrix $\boldsymbol{N}$ in (10). This scheme was originally motivated to minimize $\langle\boldsymbol{\theta}, \boldsymbol{M}\boldsymbol{\theta}\rangle$, which is called the (weighted) *algebraic distance*, hence the name "algebraic method", subject to the constraint $\langle\boldsymbol{\theta}, \boldsymbol{N}\boldsymbol{\theta}\rangle$ = constant. The solution of (10) is obtained by iteration: we first regard $W_\alpha^{(kl)}$ and $\boldsymbol{N}$ as given and solve the generalized eigenvalue problem (10), then update $W_\alpha^{(kl)}$ and $\boldsymbol{N}$ using the resulting $\boldsymbol{\theta}$, and repeat this process.

If we choose $W_\alpha^{(kl)} = 1$ and $\boldsymbol{N} = \boldsymbol{I}$ (hence no iterations are necessary), this method is nothing but the standard *least squares*. If we choose $W_\alpha^{(kl)} = 1$ and $\boldsymbol{N} = V_0[\boldsymbol{x}_\alpha]$ (the covariance matrix of $\boldsymbol{x}_\alpha$ up to scale), this reduces the method of Taubin [10], which is known to produce a fairy accurate solution. Efforts were made to improve the accuracy of the Taubin method by choosing optimal $\boldsymbol{N}$, resulting in a scheme called *HyperLS* [11], which is non-iterative. If we use the same $\boldsymbol{N}$ as in the Taubin method but use the weights $W_\alpha^{(kl)}$ that appear in (7) and (8), we obtain the iterative scheme of *renormalization* [12]. Like for HyperLS, we can optimize the matrix $\boldsymbol{N}$ of renormalization to obtain *hyper-renormalization* [13],

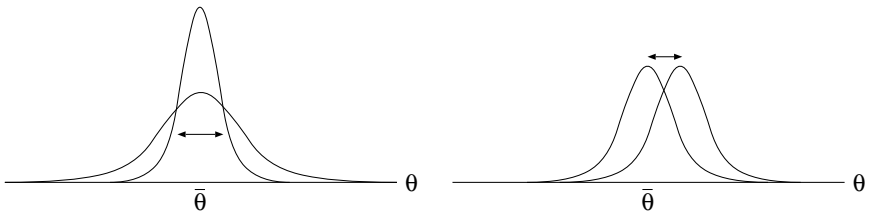

Figure 1: *Left*: The matrix $\boldsymbol{M}$ controls the covariance of $\boldsymbol{\theta}$. *Right*: The matrix $\boldsymbol{N}$ controls the bias of $\boldsymbol{\theta}$.

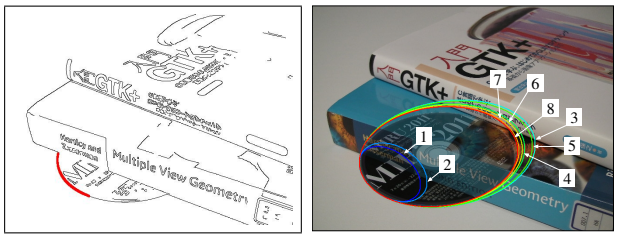

Figure 2: *Left*: An edge image of a scene with a circular object. An ellipse is fitted to the 160 edge points indicated. *Right*: Fitted ellipses superimposed on the original image. The occluded part is artificially composed for visual ease. (1) Least squares, (2) iterative reweight, (3) Taubin method, (4) renormalization, (5) HyperLS, (6) hyper-renormalization, (7) ML, (8) ML followed by hyperaccurate correction. (From [4])

which exhibits higher accuracy than maximum likelihood [5].

The superiority of hyper-renormalization is confirmed by statistical analysis. If we regard the input $\boldsymbol{x}_\alpha$ as random variables, the computed solution $\boldsymbol{\theta}$ of (10) is also a random variable. It can be shown that the matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ of (10) controle, respectively, the covariance and the bias of $\boldsymbol{\theta}$ (Fig. 1) and that the matrices $\boldsymbol{M}$ and $\boldsymbol{N}$ of hyper-renormalization are such that the covariance of $\boldsymbol{\theta}$ reaches the KCR lower bound up to $O(\sigma^4)$ and the bias of $\boldsymbol{\theta}$ is $\boldsymbol{0}$ up to $O(\sigma^4)$.

On the other hand, efforts have been made to improve the accuracy of maximum likelihood by correcting the solution a posteriori, called *hyperaccurate correction* [14]. It can be shown that maximum likelihood followed by hyperaccurate correction can achieve equivalent accuracy to hyper-renormalization [5] (Fig. 2). However, iterations for computing maximum likelihood solution sometimes fail to converge in the presence of large noise, compared to which hyper-renormalization iterations are rather robust.

References

[1] Chernov, N., Lesort, C. (2004) Statistical efficiency of curve fitting algorithms. Comput. Stat. Data Anal., 47(4), 713–728

[2] Kanatani, K. (2008) Statistical optimization for geometric fitting: Theoretical accuracy analysis and high order error analysis. Int. J. Comput. Vis., 80(2), 167–188

[3] Neyman, J., Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16(1), 1–32

[4] Kanatani, K., Sugaya, Y., Kanazawa, Y. (2016a) Ellipse Fitting for Computer Vision: Implementation and Applications. Morgan Claypool, San Rafael, CA, U.S.

[5] Kanatani, K., Sugaya, Y., Kanazawa. Y. (2016b) Guide to 3D Vision Computation: Geometric Analysis and Implementation. Springer, Cham, Switzerland

[6] Sampson, P.D. (1982) Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm. Comput. Graphics Image Process., 18(1), 97–108

[7] Chojnacki, W., Brooks, M.J., van den Hengel, A., Gawley, D. (2000) On the fitting of surfaces to data with covariances. IEEE Trans. Patt. Anal. Mach. Intell., 22(11), 1294–1303

[8] Leedan, Y., Meer, P. (2000) Heteroscedastic regression in computer vision: Problems with bilinear constraint. Int. J. Comput. Vis., 37(2), 127–150

[9] Kanatani, K., Sugaya, Y. (2010) Unified computation of strict maximum likelihood for geometric fitting. J. Math. Imaging Vis., 38(1), 1–13

[10] Taubin, G. (1991) Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. Patt. Anal. Mach. Intell., 13(11), 1115–1138

[11] Kanatani, K., Rangarajan, P., Sugaya, Y., Niitsuma, H. (2011) HyperLS for parameter estimation in geometric fitting. IPSJ Trans. Comput. Vis. Appl., 3, 80–94

[12] Kanatani, K. (1993) Renormalization for unbiased estimation. Proc. 4th Int. Conf. Comput. Vis., Berlin, 599–606

[13] Kanatani, K, Al-Sharadqah, A., Chernov, N., Sugaya, Y. (2014) Hyper-renormalization: Non-minimization approach for geometric estimation. IPSJ Trans. Comput. Vis. Appl., 6, 143–149

[14] Kanatani, K., Sugaya, Y. (2013) Hyperaccurate correction of maximum likelihood for geometric estimation. IPSJ Trans. Comp. Vis. Appl., 5, 19–29