**Annales
Geophysicae**

Least-squares multi-spacecraft gradient calculation with automatic error estimation

J. De Keyser

Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3, 1180 Brussels, Belgium

Abstract. Multi-spacecraft missions allow the gradient of important physical quantities in the terrestrial environment to be determined. The gradient can be computed from four simultaneous measurements in a straightforward way, but this computation does not produce proper error estimates, making it hard to assess the meaningfulness of the result. Recently developed least-squares gradient computation techniques offer the possibility to obtain more precise results with all-inclusive error estimates, provided that information about the non-linearity of the space and time variations of the observed quantity is given. The present paper describes several heuristics for estimating these variations, thereby enabling a fully automatic computation of the gradient and the associated error estimates. The performance of these heuristics is illustrated with synthetic data corresponding to 4- and 10-spacecraft configurations.

Keywords. Magnetospheric physics (Magnetospheric configuration and dynamics; Instruments and techniques)

1 Introduction

Computing gradients from in situ measurements is an essential element of multi-spacecraft missions, in particular the CLUSTER mission consisting of four identical spacecraft flying in formation. The classical gradient computation (CGC) technique exploits the fact that exactly four simultaneous non-coplanar measurements are needed to determine the three spatial gradient components (Harvey, 1998; Chanteur, 1998; Chanteur and Harvey, 1998; Robert et al., 1998a; Darrouzet et al., 2006). Although the basic idea is simple, such gradient computations are difficult in practice. A first set of problems has to do with the requirement of homogeneity: Computing the gradient makes sense only if

the true gradient does not deviate too much from the average gradient over the spacecraft configuration. A second set of problems is related to the measurement precision: gradients are differences of data values that differ only slightly, which inevitably leads to large relative errors on the results. This is true in particular when the spacecraft are closely spaced, which is often required by the homogeneity conditions. A third set of problems is due to imperfect knowledge of the exact place and time where the measurements are made, owing to uncertainties in the spacecraft positions, spacecraft clock synchronization errors, and the data acquisition time duration. Finally, while it is possible to assess the uncertainty on the computed spatial gradient that results from the measurement errors, the instantaneous four-spacecraft calculation provides no information about the error that stems from the fact that the gradient in reality is not constant over the spacecraft tetrahedron.

Many of these difficulties can be succesfully addressed by least-squares gradient computation (LSGC), as recently described by De Keyser et al. (2007). The rationale is that, if the gradient remains constant over a given time interval (thus relaxing the requirement of simultaneity), the information content from a larger set of data points can be exploited for computing the gradient. An overdetermined problem is then obtained from which the space-time gradient can be computed in a weighted least-squares sense. The error made by approximating the field by a locally linear one ("approximation error" or "curvature error") is related to the distance of the measurement points from the point where the gradient is computed. The approximation error is expressed in terms of the homogeneity length and time scales. The total error on the data consists of this approximation error and the measurement error. The inverse of the total error is used as the weight of the measurement, so that only points close to the center of the homogeneity domain contribute to the solution. An error estimate on the gradient is obtained that accounts for both sources of error. Many data points are usually involved in the

Correspondence to: J. De Keyser
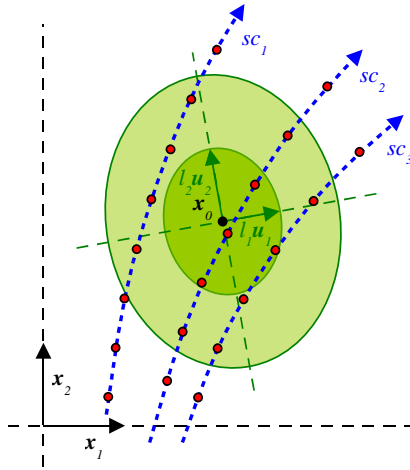(johan.dekeyser@bira-iasb.oma.be)

Fig. 1. Illustration of least-squares gradient computation in a 2-dimensional setting. The algorithm uses data obtained in a set of points in space-time. In this example, the data are acquired by three spacecraft (red dots on the dotted spacecraft trajectories); the method can deal with any number of spacecraft. The approximation error in each data point grows with its distance from x_0, where the gradient is computed. This distance is measured in a frame ($l_1 u_1$, $l_2 u_2$) that may be rotated and scaled relative to the original frame. Points on the ellipse with semi-axes l_1 and l_2 are assigned a unit distance. Points inside the ellipse (dark shaded area) correspond to smaller distances and therefore a smaller error. Points outside that ellipse (lightly shaded region) will have a larger error so that they are less relevant, thus reflecting the homogeneity condition. Points outside the shaded regions are ignored.

calculation of an individual gradient, depending on the size of the homogeneity domain relative to the spacecraft separations and the sampling frequency; the uncertainty due to random measurement errors can therefore be somewhat suppressed.

Gradient computations suffer badly from systematic errors on the data: They require properly intercalibrated data. However, intercalibration is difficult as the instruments and their operating environments are never identical. Systematic gradient computations with ESA's CLUSTER multi-spacecraft mission have been limited to magnetometer data (FGM instrument, Balogh et al., 1997, 2001) with their high precision and good calibration (Dunlop et al., 2001; Dunlop and Balogh, 2005; Vallat et al., 2005; Dunlop et al., 2006), and to electron density data derived from the plasma frequency (WHISPER instrument, Décréau et al., 1997, 2001; Trotignon et al., 2003) because of their absolute calibration (Darrouzet et al., 2006; De Keyser et al., 2007).

De Keyser et al. (2007) assume that the homogeneity length and time scales are given. While they point out that a suitable value can be chosen based on physical considerations, this may not always be easy to do in practice. They also take this scale factor to be constant over the analysis time interval, while the degree of curvature may change as

the spacecraft traverse different regions of geospace. The goal of the present paper is to introduce heuristic techniques to estimate the homogeneity scales (and thus the approximation error) automatically. Armed with such estimates, this least-squares gradient computation with adaptive scales (LSGC-AS) will use an appropriately sized homogeneity domain with the optimal set of data points, and the total error estimate on the computed gradient will be much more realistic.

Section 2 briefly reviews the least-squares gradient computation technique. We adopt standard linear algebra notation: Bold lower-case symbols represent vectors, bold upper-case symbols are matrices, and all other symbols denote scalars. Section 3 introduces various ways for modelling the approximation error. Section 4 describes techniques for automatic estimation of the parameters in those descriptions. The techniques will be illustrated with synthetic data corresponding to 4- and 10-spacecraft configurations in Sect. 5. The paper ends with an evaluation of the proposed techniques.

2 Least-squares gradient computation

This section summarizes the least-squares technique developed by De Keyser et al. (2007) for computing the gradient. We slightly generalize the description of the approximation error as this will turn out to be useful later on.

2.1 Problem formulation

Consider a scalar field $f(x, t)$ that is sampled at positions and times $x_i = [x_i; y_i; z_i; t_i]$, $i = 1, \ldots, N$, relative to a given reference frame in 4-dimensional space-time. The measurements f_i have known random error variances $\delta f_{m_i}^2$ and there are no systematic errors. The cross-correlations between measurement errors at different points vanish. To illustrate the idea, consider the 2-dimensional situation sketched in Fig. 1. We want to compute the gradient at x_0 from measurements made by several spacecraft (sc_1, sc_2, ...). The measurement points x_i are indicated by the red dots. The field f can be locally approximated by a Taylor expansion around x_0. With $\Delta x = x - x_0$ the relative position of a measurement point, and denoting the function value, the gradient, and the Hessian at x_0 by f_0, $g_0 = \nabla_{xt} f_0$, and $H_0 = \nabla_{xt} \nabla_{xt}^\top f_0$, this expansion gives

$$f(x) = f_0 + \Delta x^\top g_0 + \frac{1}{2} \Delta x^\top H_0 \Delta x + \ldots \quad (1)$$

Truncating this expansion after the linear term defines the approximating function $f_a(x) = f_0 + \Delta x^\top g_0$ and the approximation error $\delta f_a(x) = \frac{1}{2} \Delta x^\top H_0 \Delta x + \ldots$. Requiring the residuals to be zero,

$$r_i = r(x_i) = f_a(x_i) - f_i = 0, \quad (2)$$

leads to a system of N equations (one for each measurement) for f_0 and g_0. The number of unknowns, M, is 5. System (2)

is usually overdetermined ($N \gg M$) so that it cannot be satisfied exactly, but it can be solved in a least-squares sense.

Approximation (1) is valid in a region around x_0 that can be described by a hyperellipsoid in 4-dimensional space-time (the dark shaded ellipse in Fig. 1). Such an ellipsoid is uniquely specified by four mutually orthogonal unit vectors u_k, which constitute the columns of a rotation matrix $\mathbf{U} = [\ldots, u_k, \ldots]$, and by the four homogeneity length and time scales l_k, which define a scaling matrix $\mathbf{L} = \mathrm{diag}([\ldots, l_k, \ldots])$ (this notation means: the diagonal matrix with the l_k on the diagonal). Transforming the problem into a new reference frame by means of $x' = \mathbf{P} x = (\mathbf{UL})^{-1} x$, the ellipsoid becomes a hypersphere. We will often use the euclidian norm in the new frame:

$$\|\Delta x'\|^2 = \Delta x^\top \mathbf{U} \mathbf{L}^{-2} \mathbf{U}^\top \Delta x.$$

If the new frame involves only a rescaling and no rotation ($\mathbf{U} = \mathbf{I}$), this simply amounts to

$$\|\Delta x'\|^2 = \sum_k \left(\frac{\Delta x_k}{l_k}\right)^2.$$

The dark shaded ellipse in Fig. 1 corresponds to $\|\Delta x'\| \leq 1$.

Assume that an estimate for the approximation error δf_a is known. How to obtain such an estimate will be the subject of Sect. 3. At present, it suffices to remark that this error increases with $\|\Delta x'\|$. The total error consists of the measurement and the approximation error,

$$\delta f_i^2 = \delta f_{\mathrm{m}i}^2 + \delta f_{\mathrm{a}i}^2;$$

these are the diagonal elements of the total error covariance matrix $\mathbf{C}^2$. For the sake of simplicity small-scale fluctuation errors are not considered here (see De Keyser et al., 2007). We also do not address spacecraft position or timing errors.

2.2 Problem solution

The gradient is computed in a weighted least-squares sense by first multiplying system (2) with $\mathbf{C}^{-1}$. If $\mathbf{C}^2$ is diagonal, this amounts to multiplying each equation with $w_i = 1/\delta f_i$, the weight for the i-th equation. If $\mathbf{C}^2$ is not diagonal, a diagonalization has to be computed first, which we will try to avoid as this can be very compute-intensive. Solving the weighted overdetermined system

$$r_i / \delta f_i = 0 \tag{3}$$

is equivalent to minimizing the least-squares expression

$$\chi^2 = \sum_{i=1}^N \frac{r_i^2}{\delta f_i^2}.$$

The choice of weights makes sure that measurements with a large total error do not contribute much to the solution. In particular, data outside the homogeneity domain have a large approximation error and do not add any significant information. To limit the amount of computational work, the set of

data points that are used is limited to those whose total error is less than a given factor σ (typically, $\sigma = 1000$) times the smallest total error in the set of data points. If the measurement errors are all equal, the domain from which data points are accepted then is a large ellipsoid (the lightly shaded ellipse in Fig. 1). The number of measurements, N, is therefore rather large, so that $N \gg M$.

Solving the weighted least-squares minimization problem is a classical application of the singular value decomposition. With $q = [f_0; \mathbf{LU}^\top g_0]$ grouping the unknowns, the linear approximation can be written as $f_a = \mathbf{A} q = [\mathbf{1}, \mathbf{L}^{-1} \mathbf{U}^\top \Delta \mathbf{X}^\top] q$. The solution of the weighted overdetermined system (3) can be expressed as

$$q = (\mathbf{A}^\top \mathbf{C}^{-2} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{C}^{-2} f = \mathbf{M} f. \tag{4}$$

The solution is obtained by applying the linear operator $\mathbf{M}$ to the data f. The inverse of the symmetrized weighted system matrix $\mathbf{A}^\top \mathbf{C}^{-2} \mathbf{A}$ is computed by means of the singular value decomposition of $\mathbf{Z} = \mathbf{C}^{-1} \mathbf{A}$ (see De Keyser et al., 2007, for more details). Expression (4) indicates how measurement errors δf produce an error on the result:

$$\delta q = \mathbf{M}\, \delta f,$$

with covariances $\mathbf{C}_q^2 = \mathbf{M} \mathbf{C}^2 \mathbf{M}^\top = (\mathbf{A}^\top \mathbf{C}^{-2} \mathbf{A})^{-1}$. Small singular values therefore imply strong error propagation. In particular, zero singular values indicate that the measurements do not completely define the solution. The singular values offer a convenient generalization of the tetrahedron geometric factors (size, elongation, and planarity) for 4-spacecraft configurations (Robert et al., 1998b). In general, $\mathbf{C}_q^2$ is not diagonal: The errors on different solution components are correlated. Such correlations are often ignored, and one adopts $[\mathrm{diag}(\mathbf{C}_q^2)]^{1/2}$ as the error margins.

The least-squares gradient method can be applied to vector fields as well; the number of unknowns then is $M = 15$. The cross-correlations between the errors on the gradients of the different field components must be taken into account when computing the error margins on the curl or the divergence of the vector field. The least-squares method can also handle constraints, e.g., when a vector field is divergence-free (De Keyser et al., 2007), in which case $M = 14$. Applied to the magnetic field, this leads to a new curlometer since $j = \nabla \times B / \mu_0$ if there is no time dependence.

The method described here relies on an unconstrained approximation of the scalar or vector field f. If f is strictly positive, like for plasma densities or temperatures, the method should be applied to $\log f$ rather than to f itself.

3 Modelling the approximation error

We consider different descriptions of the approximation error δf_a, depending on the desired level of detail.

Table 1. Relative contributions of different regions in $\Delta \boldsymbol{x}'$-space to the least-squares solution for different approximation error models, assuming uniform coverage. See the text for an interpretation.

d	$\Delta x' \in [0, \frac{1}{2}]$	$\Delta x' \in [\frac{1}{2}, 1]$	$\Delta x' \in [1, 2]$	$\Delta x' > 2$
(A)	$\delta f_a \propto \Delta x'^2$			
1	0.4447	0.3358	0.1829	0.0365
2	0.1560	0.3440	0.3440	0.1560
3	0.0365	0.1829	0.3359	0.4447
4	0.0016	0.0172	0.0581	0.9231
(B)	$\delta f_a \propto \Delta x'^2 \max\{\Delta x'^2, 1\}$			
1	0.5100	0.3851	0.1037	0.0012
2	0.2380	0.5251	0.2318	0.0051
3	0.1028	0.5144	0.3670	0.0158
4	0.0410	0.4278	0.4890	0.0422
(C)	$\delta f_a \propto \Delta x'^2 \max\{e^{\Delta x'-1}, 1\}$			
1	0.6647	0.3031	0.0317	0.0004
2	0.4009	0.5034	0.0934	0.0023
3	0.2181	0.5831	0.1899	0.0089
4	0.1058	0.5529	0.3138	0.0275
(D)	$\delta f_a \propto \Delta x'^2 \max\{e^{\Delta x'^2-1}, 1\}$			
1	0.6450	0.3344	0.0206	0.0000
2	0.3892	0.5542	0.0566	0.0000
3	0.2209	0.6684	0.1106	0.0001
4	0.1184	0.6980	0.1835	0.0002

3.1 Weak and strong locality

It is clear that the approximation tends to degrade with distance $\Delta x' = \|\Delta \boldsymbol{x}'\|$ away from $\boldsymbol{x}_0$, where the gradient has to be computed. By choosing the weights inversely proportional to the total error, the "(weak) locality principle" is satisfied: an individual point closer to $\boldsymbol{x}_0$ contributes more to the least-squares solution than a point far away. De Keyser et al. (2007) use this principle to argue that the set of points included in the calculation can be limited (cf. the σ threshold that was introduced in Sect. 2). Upon closer inspection, however, it turns out that one must be very careful in drawing that conclusion. Let us assume that the approximation error can be modelled as $\delta f_a \le f_c \Delta x'^p$, where f_c is a scaling constant reflecting the approximation error at $\Delta x'=1$. If the measurement errors are all the same, and equal to f_c, the weights used in the least-squares problem are

$$w_i^2 = \frac{1}{\delta f_i^2} = \frac{1/f_c^2}{1 + \Delta x'^{2p}}.$$

The contribution ζ of the set of points inside the homogeneity domain to the computed solution can be quantified by summing the weights of these points, and comparing that to

the sum of the weights of all points. If the data points are distributed uniformly in d-dimensional space, and if this set of points is dense enough to use the continuous-distribution limit, it is found that

$$\zeta = \lim_{R \to \infty} I_d(1)/I_d(R),$$

where

$$I_d(R) = \int_0^R w(\rho) S_d(\rho)\, d\rho = \int_0^R \frac{dC_d \rho^{d-1}}{1 + \rho^{2p}}\, d\rho,$$

in which the constant $C_d = V_d(\rho)/\rho^d$, with $V_d(\rho)$ and $S_d(\rho)$ representing the volume and surface of a d-dimensional sphere with radius ρ, respectively. Whatever the value of p, there is always a dimension d for which I_d is unbounded for $R \to \infty$, so that $\zeta \to 0$: The far-away points dominate the solution despite the fact that the weak locality principle is satisfied. Requiring that the solution is not dominated by the set of points outside the homogeneity domain, which we refer to as the "strong locality principle", leads to $2p > d-1$. For the space-time setting of the multi-spacecraft gradient problem ($d \le 4$), this requirement is satisfied for $p \le 2$. To be general for all d, δf_a must increase at least exponentially with $\Delta x'$: Only if the approximation error increases rapidly, the sum of the weights of far-away points decreases quickly enough to yield a finite contribution. Table 1 lists the relative contributions of the sets of points with different $\Delta x'$-ranges in the uniform and continuous distribution limit, for four approximation error bounds. For $d=4$, a bound $\delta f_a \propto \Delta x'^2$ is problematic as more than 90% of the solution is contributed by points outside $\Delta x' > 2$. A higher p or an exponential error bound is needed to enforce strong locality; the table gives a few alternatives. The assumption of a uniform coverage of the homogeneity domain is often not realistic, as the number of spacecraft is usually limited. As an alternative, consider the case where the data cover a d-dimensional cylinder whose cross-section has dimensions well below the homogeneity scales, as for closely spaced spacecraft crossing a large structure. The contribution then is proportional to I_1, giving the results listed in Table 1 for $d=1$, regardless of the actual dimension d. For the quadratic approximation error bound, the contribution from points outside $\Delta x' > 2$ is less than 4%, so that the situation is not that bad. Note that these assessments of ζ in different circumstances are rather crude since the approximation errors at different points are not statistically independent. Summing or integrating the squared weights is therefore only indicative of the contribution of a particular set of points to the solution.

3.2 Homogeneity properties and the second-order term

The approximation error is due to the second- and higher-order terms in the Taylor expansion (1):

$$\delta f_a(\Delta \boldsymbol{x}') = \delta f_{so}(\Delta \boldsymbol{x}') + \delta f_{ho}(\Delta \boldsymbol{x}'),$$

where $\delta f_{\rm so}=\mathcal{O}(\Delta x'^2)$ is the second-order term, corresponding to the term with the Hessian. It can be written as

$$\delta f_{\rm soi} = \frac{1}{2}\boldsymbol{\Delta x}_i{}^\top \mathbf{H}_0 \boldsymbol{\Delta x}_i = f_c \sum_{k=1}^{d} s_k \Delta x'^2_{ik}$$

if the Hessian is $\mathbf{H}_0=\mathbf{U\Lambda U}^\top$ with $\mathbf{\Lambda} = \mathrm{diag}(2f_c s_k/l_k^2)$, that is, if the eigen-vectors of the Hessian are the homogeneity directions, if its eigen-values λ_k determine the homogeneity lengths $l_k=\sqrt{2f_c/|\lambda_k|}$ and if the sense of curvature $s_k=\mathrm{sign}\lambda_k$ in each direction, with f_c an a priori given scaling constant. The homogeneity properties therefore determine the second-order term in the Taylor approximation.

3.3 Description of the approximation error

Because of the strong locality principle, we use

$$\delta f_{\rm ai} = f_c \sum_{k=1}^{d} s_k \Delta x'^2_{ik} \cdot \max\{\eta(\boldsymbol{\Delta x}')\phi(\Delta x'), 1\}$$

to express the approximation error, where $\eta(\boldsymbol{\Delta x}')\in[-1, 1]$ is bounded and $\phi(\Delta x')$ is a monotonically increasing function with $\phi(1)=1$. Four approximation error models are considered here: (A) $\phi\equiv 1$, (B) $\phi=\Delta x'^2$, (C) $\phi=\mathrm{e}^{\Delta x'-1}$, and (D) $\phi=\mathrm{e}^{\Delta x'^2-1}$. The actual behaviour of function $\eta(\boldsymbol{\Delta x}')$, and especially its sign, is not known. This limits our ability to estimate the variances. A simple estimate is

$$\langle\delta f_{\rm a}^2{}_i\rangle \leq f_c^2 (\sum_{k=1}^{d} s_k \Delta x'^2_{ik})^2 \max\{[\phi(\Delta x')]^2, 1\}.$$

For $\Delta x'=1$, the variance is $\delta f_{\rm a}^2 \leq f_c^2$, in line with the definition of the homogeneity lengths as the scales corresponding to an approximation error f_c. The above approximation error estimate is correct for small $\Delta x'$ when the homogeneity properties are well estimated. The higher-order terms are described rather poorly, but points where those terms play a role do not contribute much to the solution anyway because of the strong locality principle. This can be appreciated by looking at the upper bounds given by the four approximation error models. These are the four alternatives discussed in Table 1; they are illustrated for the one-dimensional case in Fig. 2. For $\Delta x'<1$ the approximation error is quadratic in all four cases. For points that are farther away, the error grows more rapidly with alternatives B, C, and D, so that such points have substantially less weight. While model B initially grows faster than model C, the converse is true as $\Delta x'\to\infty$. Model D produces a very steep increase of the approximation error, so that points beyond roughly $\Delta x'>1.7$ do not count at all. In the sequel, model C has been adopted, as it offers the best compromise between guaranteeing strong locality and allowing points outside $\Delta x'=1$ to contribute some information to the solution.

The approximation errors at two points i and j are not independent. Because of the lack of knowledge about η, we
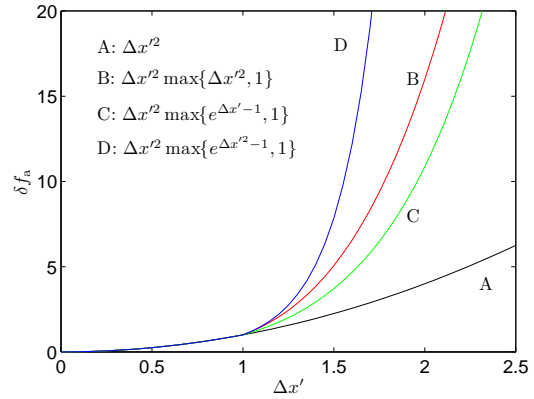

Fig. 2. Behaviour of four different upper bounds for the approximation error. These four models adopt a quadratic behaviour for $\Delta x'<1$, but differ in the way in which they estimate the higher-order terms in the Taylor approximation for larger $\Delta x'$.

can only assume that the covariances of the higher-order part of the solution vanish. In fact, as the higher-order terms represent a residual error, typically with oscillatory behaviour, such correlations can indeed often be ignored. The covariances therefore are due to the second-order part alone:

$$\langle\delta f_{\rm ai}\delta f_{\rm aj}\rangle \approx f_c^2 (\sum_{k=1}^{d} s_k \Delta x'^2_{ik})(\sum_{k=1}^{d} s_k \Delta x'^2_{jk}).$$

In the particular situation where the signs s_k are not known, an upper bound on the variances is given by

$$\delta f_{\rm ai}^2 \approx f_c^2 \Delta x'^4 \{[\phi(\Delta x')]^2, 1\};$$

the covariances can only be taken zero. This simplified description with a diagonal covariance matrix $\mathbf{C}^2$ is computationally much cheaper.

3.4 Role of error cross-correlations

Consider the following toy problem for $d=1$ with $s_1=+1$, where the data points can be grouped in three sets: set I containing points well inside the homogeneity domain ($\Delta x'\ll 1$), set II grouping points near the edge of the homogeneity domain ($\Delta x'\approx 1$), and set III containing points outside the homogeneity domain ($\Delta x'=\xi>1$). These sets contain N_a, N_b, and N_c points, respectively. For the sake of simplicity, consider $\delta f_{\rm mi}\equiv f_c$. The covariance matrix then is

$$\mathbf{C}^2 = f_c^2 \begin{bmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}+\boldsymbol{E} & \xi^2\boldsymbol{E} \\ \mathbf{0} & \xi^2\boldsymbol{E} & [1+\xi^4([\phi(\xi)]^2-1)]\boldsymbol{I}+\xi^4\boldsymbol{E} \end{bmatrix},$$

with cross-correlations due to the quadratic terms ($\boldsymbol{I}$ denotes identity matrices, $\boldsymbol{E}$ represents matrices whose elements are

all 1). With a scaling $\mathbf{S}_C = f_c \text{diag}([\ldots 1 \ldots \sqrt{2} \ldots s \ldots])$, where $s^2 = 1 + \xi^4 [\phi(\xi)]^2$, one obtains

$$\mathbf{C}^2 = \mathbf{S}_C^\top \begin{bmatrix} I & 0 & 0 \\ 0 & \frac{1}{2}(I+E) & \frac{\xi^2}{\sqrt{2}s}E \\ 0 & \frac{\xi^2}{\sqrt{2}s}E & (1-\frac{\xi^4}{s^2})I + \frac{\xi^4}{s^2}E \end{bmatrix} \mathbf{S}_C$$

If ϕ is strictly increasing (as in models B, C, and D), the cross-correlations between points of sets II and III, and those among points of set III, are $1/\sqrt{2}\phi(\xi)$ and $1/[\phi(\xi)]^2$, which tend to zero as $\xi \to \infty$, so that

$$\mathbf{C}^2 \approx \mathbf{S}_C^\top \begin{bmatrix} I & 0 & 0 \\ 0 & \frac{1}{2}(I+E) & 0 \\ 0 & 0 & I \end{bmatrix} \mathbf{S}_C.$$

This indicates that only total error cross-correlations between points of set II matter: Those involving points of set I are negligble since uncorrelated measurement errors dominate the total error there, while those involving points of set III are negligible since the second-order error is dwarfed by the uncorrelated higher-order contributions there. The approximate eigen-values of $\mathbf{C}^2$ are the variances of set I, N_a times f_c^2, the variances of set III, N_c times $(1+\xi^4[\phi(\xi)]^2)f_c^2$, and those of set II, which for the special form of the $\frac{1}{2}(I+E)$ diagonal block are known to be once $(N_b-1)f_c^2$ and N_b-1 times f_c^2. The corresponding eigen-vectors indicate how the original set of equations is reordered into an equivalent set of equations that each represent a statistically independent piece of information. The equations for set I remain unaffected; the eigen-values indicate essentially the independent measurement error variances. The equations for set III are unaffected as well, corresponding to variances that reflect the higher-order errors. The equations for the points of set II, however, are linearly combined into a set of N_b-1 equations with an improved precision as the correlated second-order part of the error can be eliminated there, while there remains one equation with a much higher variance, which is equivalent to dropping that equation from the system.

Generalizing the conclusions of this toy problem to arbitrary distributions of data points, as well as to the multi-dimensional case, one finds that adding the cross-correlations helps to partially eliminate the effect of the approximation error in points near and beyond the edge of the homogeneity domain, but the contribution of those points is limited because of the strong locality principle. Overall, including the correlation tends to reduce the error margin on the gradient a little, but it does not have a dramatic effect. Even rough estimates of the cross-correlations are therefore sufficient.

This has important practical consequences. Solving the weighted overdetermined problem requires inverting the total error covariance matrix. This computationally very expensive operation can be accelerated, for instance, by setting small correlations to zero in order to improve the sparsity of the matrix. An even more dramatic acceleration is obtained by simply ignoring the cross-correlations, thus avoiding the inversion of the correlation matrix altogether.

4 Automatic determination of homogeneity parameters

The LSGC technique explained in Sect. 2 requires that an approximation error estimate is given. Section 3 showed how that error can be described in terms of the homogeneity directions $\boldsymbol{u}_k$, the homogeneity scales l_k, and the curvature senses s_k. Applying the least-squares gradient method would be much easier if these parameters could be determined automatically. It can intuitively be understood that this must somehow be possible: the residuals reflect the behaviour of the error, from which the parameter values can be extracted.

We assume here that the orientations $\boldsymbol{u}_k$ in space-time are given: Most often, one direction is along the time axis, a second one is along the magnetic field direction, while there is no a priori homogeneity anisotropy perpendicular to the field. The set of parameter values that must be estimated therefore is either $\mathcal{S} = \{l_k\}$, or $\mathcal{S} = \{l_k, s_k\}$, depending on the desired level of detail. In all cases, approximation error model C has been used. It can sometimes be useful to adopt a more complicated choice for the $\boldsymbol{u}_k$. For instance, for convecting time-stationary structures, and knowing the convection speed (e.g. from plasma measurements), it can be advantageous to consider computing the gradients in the comoving frame defined by the specific choice of the $\boldsymbol{u}_k$ as discussed by De Keyser et al. (2007, Appendix A2).

We will determine the parameter values by optimization: The least-squares gradient is computed for different sets of parameter values so as to minimize a function $\mathcal{F}(\mathcal{S})$ that represents the quality of a set. Different functions $\mathcal{F}$ lead to different heuristic parameter estimation techniques.

4.1 Heuristics based on χ^2

If $\boldsymbol{q}$ denotes the gradient computed with parameter values $\mathcal{S}$, then $\mathbf{A}\boldsymbol{q}$ are the observations that would have been made if the gradient were exact, and $\boldsymbol{r} = \mathbf{A}\boldsymbol{q} - \boldsymbol{f}$ are the residuals. It is a basic property of χ^2-statistics that

$$\chi^2 = \boldsymbol{r}^\top \mathbf{C}^{-2} \boldsymbol{r} = \sum_{i=1}^N \left(\frac{r_i}{\delta f_i}\right)^2 = N - M, \tag{5}$$

the effective number of degrees of freedom. We can use this property to estimate the approximation error.

Starting with a set of parameter values $\mathcal{S}^{(n-1)}$, the gradient is computed, as well as

$$\chi^2 = \sum_i r_i^{(n-1)2} / \delta f_i^{(n-1)2}.$$

If this value matches $N-M$, the specified variances did correspond to the observed variability. If this is not the case, something was wrong with the given error estimates $\delta f_i^{(n-1)}$.

If the measurement errors are well-known, the approximation error estimates must have been incorrect. Improved estimates $\delta f_i^{(n)}$ and the corresponding residuals $r_i^{(n)}$ should satisfy

$$N - M = \sum r_i^{(n)2}/\delta f_i^{(n)2}.$$

Changing the weights in the overdetermined problem only has a modest effect on its solution, so that $r_i^{(n)} \approx r_i^{(n-1)}$ and

$$\delta f_i^{(n)2} \approx \frac{\chi^2}{N-M}\delta f_i^{(n-1)2} = \alpha^{(n)2}\delta f_i^{(n-1)2};$$

the factor $\alpha^{(n)}$ should be used to rescale the total error estimates. To understand the consequences of this, assume that we are working with approximation error model A (second-order term only, without knowledge of curvature senses) and that the homogeneity directions are along the coordinate axes. Denoting the improved homogeneity lengths by $l_k^{(n)} = \lambda^{(n)} l_k^{(n-1)}$, the total error estimates are

$$\delta f_i^{(n-1)2} = \delta f_{mi}^2 + \left[\sum_k (\frac{\Delta x_{ik}}{l_k^{(n-1)}})^2\right]^2,$$

$$\delta f_i^{(n)2} = \delta f_{mi}^2 + \left[\sum_k (\frac{\Delta x_{ik}}{\lambda^{(n)}l_k^{(n-1)}})^2\right]^2,$$

so that in an average sense

$$\lambda^{(n)-4} = \alpha^{(n)2} + (\alpha^{(n)2}-1)\frac{\langle\delta f_{mi}^2\rangle}{f_c^2}/\left\langle\left[\sum_k \left(\frac{\Delta x_{ik}}{l_k^{(n-1)}}\right)^2\right]^2\right\rangle.$$

Such a solution exists only if the right hand side is positive. This condition is satisfied if $\alpha^{(n)} \geq 1$, when the approximation error was underestimated. A solution also exists if $\alpha^{(n)} < 1$, but only if the measurement error is small compared to the approximation error: If the measurement errors dominate, rescaling the homogeneity lengths does not affect the residuals very much. In the particular situation where $\langle\delta f_{mi}^2\rangle \ll f_c^2$, this amounts to rescaling the homogeneity lengths with a factor $\lambda^{(n)} = 1/\sqrt{\alpha^{(n)}}$.

In conclusion: If the obtained χ^2 was too large, the approximation error was underestimated and the homogeneity lengths must be decreased. If χ^2 was too small, the converse is true. Because of the simplifying assumptions made above, this adaptation process is an iterative one. As the process is repeated, $\alpha^{(n)} \to 1$. This amounts to minimizing

$$\mathcal{F} = \alpha^2 + \frac{1}{\alpha^2} \geq 2,$$

a technique that we call "local χ^2 optimization". The minimum of $\mathcal{F}$ is uniquely defined if the curvature lengths are $l_k = (\lim_{n\to\infty} \lambda^{(n)} \cdot \ldots \cdot \lambda^{(1)})l_k^{(0)} = \lambda l_k^{(0)}$ with a direction-independent proportionality constant, so that minimizing $\mathcal{F}(l_k(\lambda))$ is a one-dimensional optimization problem. The

heuristic does not allow to estimate the individual homogeneity lengths, nor the sense of curvature. Different initial homogeneity scales in various directions can be specified; they are rescaled while keeping their relative proportions.

The value of $\alpha^{(n)}$ and the corresponding scale $\lambda^{(n)}$ are some sort of mean value as the technique cannot distinguish the relative contributions from different directions. If one is dealing with gradients of one-dimensional structures (a rather common situation), the approximation error is due to one of the d dimensions only, so that the actual rescaling factor must be taken d times larger. Even if there is no single dominant curvature direction, it is safe to do so. Also, as we work with three-standard-deviation error bounds rather than the one-standard-deviation bounds used in the defining property of χ^2, the corresponding factor must be added. Finally, in order to give more emphasis to points where the approximation error is small, we introduce an additional factor $\mu = 2$ to force the homogeneity lengths to be chosen a little bit smaller than they would otherwise. We therefore define

$$\alpha^{(n)2} = 9\mu^2 d^2 \frac{\chi^2}{N-M},$$

producing a rescaling that reflects the smallest spatial scale.

If the number of available data, N, is not large, the statistical properties of χ^2 cannot be relied upon. This problem can be overcome by determining the correction not on the basis of an individual gradient computation, but for a set of gradient computations performed over a certain time interval. Doing so leads to proportionally larger values of N, M, and $N-M$, so that χ^2-statistics do apply. The homogeneity scales l_k then remain constant throughout the time interval, so that this approach is useful only when the curvature properties do not change significantly over that interval. We refer to this technique as "global χ^2 optimization".

4.2 Heuristics based on the distribution of the residuals

More detailed heuristics analyze the spatial distribution of the residuals to find the l_k in each direction individually. Also the curvature signs s_k can be determined.

Figure 3 sketches a typical distribution of the (non-weighted) residuals for the one-dimensional case. The measurement errors lead to non-zero values for $\|\Delta x\|$ close to zero. Farther away, the second-order terms lead to a quadratic behaviour. Still farther away, the higher-order terms dominate so that there is not necessarily an obvious systematic behaviour anymore. We can subtract the measurement errors and estimate the approximation errors

$$\delta f_{ai}^2 \approx \max\{r_i^2 - 3\langle\delta f_{mi}^2\rangle, 0\},$$

from which we try to recover the properties of the second-order errors.

A first possibility is to fit the approximation errors with a second-order expression of the form

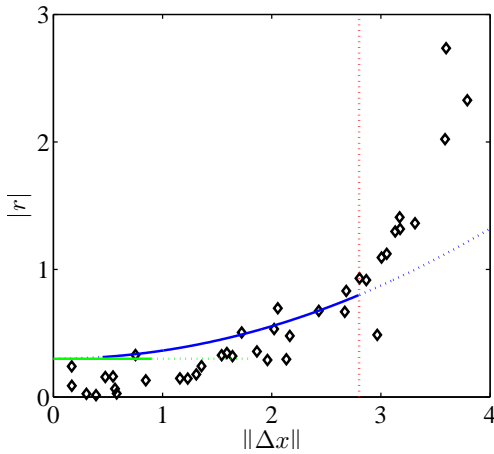$$|\delta f_{ai}| = \epsilon f_c \Delta x_i'^2$$

Fig. 3. Typical spatial distribution of the residuals in the one-dimensional case. The diamonds indicate the absolute values of the residuals. For small $\|\Delta x\|$, the measurement errors dominate (measurement error bound: green line). At intermediate values, the second-order behaviour is evident (measurement error plus second-order term: blue curve). For larger $\|\Delta x\|$, higher-order terms play a role and may result in residuals whose upper bound can be higher or lower. The goal of solution adaptivity is to find the homogeneity length scale (marked by the dashed vertical line), at which the transition of second-order to higher-order behaviour occurs.

to determine the parameter $\epsilon > 0$. This is an overdetermined problem. To eliminate the impact of the higher-order terms, weights $w_i^2 = e^{-\Delta x_i'^2/X^2}/\Delta x_i'^2$ are associated with each equation; $X = 3$ is used here. The weighted least-squares solution is

$$\epsilon = \frac{\sum_i e^{-\Delta x_i'^2/X^2} |\delta f_{ai}|}{f_c \sum_i e^{-\Delta x_i'^2/X^2} \Delta x_i'^2}.$$

We take the value 3ϵ to be an approximate upper bound for the approximation error. It is again possible to obtain a suitable value corresponding to the smallest homogeneity scale by setting $\alpha^{(n)} = 3\mu d\epsilon$, similar to what was done for χ^2 optimization, implying a rescaling of the homogeneity lengths by a common factor $\lambda^{(n)} = 1/\sqrt{\alpha^{(n)}}$ at each step. One can again define an optimization process that determines the homogeneity scales that minimize

$$\mathcal{F} = \alpha^2 + \frac{1}{\alpha^2} \geq 2.$$

As $\alpha^{(n)} \to 1$ near the optimum, the homogeneity lengths $l_k = (\lim_{n\to\infty} \lambda^{(n)} \cdot \ldots \cdot \lambda^{(1)}) l_k^{(0)} = \lambda l_k^{(0)}$ have the desired values. This "solution-adaptive technique with common rescaling" is similar to local χ^2 optimization.

A direction-dependent fit of the form
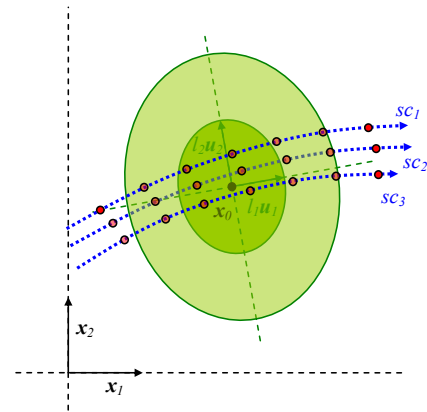
$$|\delta f_{ai}| = f_c \sum_k \epsilon_k \Delta x_{ik}'^2$$

Fig. 4. If the spacecraft sample the homogeneity domain only partially, for instance only along homogeneity direction $\boldsymbol{u}_1$ as sketched in this example, the residuals do not contain any information about the approximation error along the other directions so that the corresponding homogeneity lengths cannot be derived.

can be used to find the parameters $\epsilon_k > 0$; this coincides with the previous technique if all $\epsilon_k \equiv \epsilon$ are identical. The weighted least-squares solution can be formally written as

$$[\epsilon_k] = \left[\sum_m w_m^2 \Delta x_{im}'^2 \Delta x_{jm}'^2\right]^{-1} \left[\sum_m w_m^2 \Delta x_{km}'^2 |\delta f_{ak}|/f_c\right]$$

where the same weighting is used as before. One has to be careful to verify that all $\epsilon_k \leq 0$; if not, one has to explore a number of combinations in which one or more of the ϵ_k are zero and the remaining values are computed from the above system of equations, and retain the solution that minimizes the weighted residuals. The effectively used values are taken as $\alpha_k^{(n)} = 3\mu\epsilon_k > 0$, from which $\lambda_k^{(n)} = 1/[\alpha_k^{(n)}]^{1/2}$ and $l_k = (\lim_{k\to\infty} \lambda_k^{(n)} \cdot \ldots \cdot \lambda_k^{(1)}) l_k^{(0)} = \lambda_k l_k^{(0)}$.

It has been our experience that this multi-dimensional fitting procedure fails to work properly if the homogeneity domain is only partially sampled. And this, unfortunately, is very often the case with data recorded by spacecraft that fly in formation along a common trajectory. Suppose that the spacecraft orbit is along one of the homogeneity directions, as depicted in Fig. 4. If the homogeneity scales in the perpendicular directions happen to be much larger than the transverse spacecraft separations, the residuals do not contain much information about the approximation error in those directions, so that the corresponding homogeneity lengths cannot be properly estimated. This is not really a problem: As there are no data points in those directions, there is no need to evaluate the corresponding approximation errors, and so those homogeneity lengths are not needed. One only has to make sure that the heuristic technique is robust enough to handle such circumstances.

The strategy adopted here is to guide the direction-dependent heuristic by the direction-independent one. First

obtain $\alpha_{\max}=1/\lambda_{\min}^2$ from the common rescaling method, corresponding to the minimum scale length in any direction. Then apply the direction-dependent technique. In each step we obtain the $\alpha_k^{(n)}$ values. Defining $\hat{\alpha}_k=\alpha_{\max}/A_k$ with

$$A_k = \alpha_k^{(n-1)} \cdot \ldots \cdot \alpha_k^{(1)},$$

we set the effective direction-dependent rescaling factors to

$$\log \alpha_{\mathrm{eff}k}^{(n)} = \log \hat{\alpha}_k + \max\{\log L^2 \cdot \mathrm{erf}\frac{\sqrt{\pi}\,\log(\alpha_k^{(n)}/\hat{\alpha}_k)}{2\log L^2}, 0\}$$

where $L>1$ is a given constant. This restricts $\alpha_{\mathrm{eff}k}^{(n)}$ to the interval $[\hat{\alpha}_k/L^2, \hat{\alpha}_k]$ if it was outside, while keeping $\alpha_{\mathrm{eff}k}^{(n)} \approx \alpha_k^{(n)}$ inside that interval. This choice guarantees that

$$\frac{1}{L^2} \le \frac{\alpha_{\mathrm{eff}k}^{(n)} A_k}{\alpha_{\max}} \le 1 \text{ and thus } 1 \le \frac{l_k^{(n)}}{\lambda_{\min} l_k^{(0)}} \le L.$$

This choice therefore makes sure that the homogeneity lengths all lie in a bounded interval above the minimum length as determined by common rescaling. This makes this "solution-adaptive technique with direction-dependent rescaling" very robust, while at the same time allowing some adaptivity. The constant L can be freely chosen, but it should not be too large to avoid ill-defined values when the homogeneity domain is not well covered. The choice $L=10$ has been adopted here, which is sufficient to capture the effects of direction-dependent homogeneity properties. The target function that is minimized in the direction-dependent case is

$$\mathcal{F} = \sum_{k=1}^{d} \alpha_k^2 + \frac{1}{\alpha_k^2} \ge 2,$$

under the constraints discussed above.

Yet another way to improve the well-posedness of the direction-dependent adaptation process is to limit the number of unknown homogeneity scales by introducing constraints. For instance, one might require the spatial homogeneity lengths to be all equal. This is done by taking equal reference lengths $l_x^{(0)}=l_y^{(0)}=l_z^{(0)}$ and by forcing $\lambda_x=\lambda_y=\lambda_z$, while the homogeneity time is determined by a reference time $l_t^{(0)}$ and a rescaling factor λ_t. The corresponding direction-dependent technique then becomes a two-dimensional optimization process, which is also computationally easier to solve.

The "solution-adaptive technique with direction-dependent rescaling and curvature" goes one step further: from the signs of the (non-weighted) residuals $r=f-\mathbf{A}q$ one can infer the curvature signs s_k. In this case, too, it can be useful to set constraints on the set of homogeneity scales; introducing constraints on the curvature signs, however, is of little practical use. The curvature sign in a particular direction is deemed to be zero if the corresponding homogeneity length is large: in such directions the curvature error is small. As explained in Sect. 3.3, knowing the signs allows a more precise evaluation of the approximation error.

It also allows the cross-correlations on the data used in the overdetermined system to be estimated, although we do not exploit that in order limit computation time.

4.3 Homogeneity parameters from a second-order fit

An alternative approach would be to compute both a first- and a second-order fit, and use the difference between both as an error estimate. Computing a second-order fit is, however, usually not feasible. A full quadratic fit implies many unknowns: $M=15$ for a scalar field (function value, gradient, and symmetric hessian matrix) and $M=45$ for a vector field. In a simplified version where the homogeneity directions are good approximations of the eigen-vectors of the hessian, the hessian off-diagonals vanish, so that $M=9$ or $M=27$ for a scalar or a vector field. In either case, there usually is not enough information available to determine all these unknowns because the homogeneity domain is not fully covered by the sampling points.

4.4 Optimization procedure

The heuristic techniques described above rely on minimizing target function $\mathcal{F}$, a multi-dimensional optimization problem. The optimization algorithm used here is the classical Broyden, Fletcher, Goldfarb, and Shanno (BFGS) method with numerical derivatives. This algorithm progressively evolves from a steepest descent method to a quasi-Newton method. Starting from an initial guess and an estimate of the target function gradient at that point (obtained by numerical differentiation), a line search is performed to find the (approximate) minimum in that direction. The line search iteratively uses an interpolating parabola to approach that minimum. This procedure is repeated, but at each step the target function gradient is used to improve an estimate of the Hessian $\mathbf{H}_{\mathcal{F}}$ of the target function near the minimum, from which the next search direction is computed. Rather than storing the Hessian, one stores its Cholesky factor $\mathbf{L}_{\mathcal{F}}$ (where $\mathbf{H}_{\mathcal{F}}=\mathbf{L}_{\mathcal{F}}^{\top}\mathbf{L}_{\mathcal{F}}$); the BFGS algorithm relies on a specific update and downdate of the Cholesky factor that is computationally efficient.

Given the algorithm to compute the least-squares gradient and the choices made therein (such as the selection of data points to be used), the target function $\mathcal{F}$, while having a smooth overall behaviour as a function of its arguments, is not necessarily smooth locally. This behaviour manifests itself when solving the optimization problem with high precision: The optimization might get trapped in a local minimum close to the global minimum. To avoid this, we let BFGS optimization be followed by a comparison of the BFGS solution with the target function values in a set of nearby alternatives in all directions. If a lower value is found, BFGS is used to improve that solution even further.

We formally describe $\mathcal{F}$ as being a function of $\log \alpha=-2\log l$ or $\log \alpha_k=-2\log l_k$. We solve each

optimization problem until the solution updates become smaller than a specified precision of 10^{-6}. This implies that the values of l or l_k are determined with a relative error of 0.1%, which is more than adequate for our purpose.

The global χ^2 optimization heuristic requires solving one optimization problem in just a single variable. On one hand, the evaluation of the target function is pretty expensive, as it consists of computing the gradients at all points for the current values of the homogeneity lengths to evaluate α. On the other hand, for a one-dimensional search space the BFGS technique reduces to a line search, so that only a limited number of evaluations is needed.

For the local χ^2 technique, an optimization problem must be solved at each point where the gradient is to be computed. Each optimization problem is one-dimensional, with a target function that involves computing one gradient, which is fairly easy.

The solution-adaptive technique with common rescaling is similar to local χ^2 optimization: Only the definition of α is different.

The solution-adaptive technique with direction-dependent rescaling is more expensive, in that there are now d or $3d$ parameters at each point, the homogeneity scales for the scalar and vector cases, respectively. (The number of parameters is less when constraints on the homogeneity scales are being used.) The heuristic benefits from the computational efficiency of BFGS for such multi-dimensional problems. We perform the optimization in three steps. First, the common rescaling problem is solved. The multi-dimensional direction-dependent problem is then solved with the BFGS algorithm, while forcing the homogeneity scales in a band within a factor $L=\sqrt{10}$ above the direction-independent minimum length. Finally, this solution is used as the starting guess for solving the direction-dependent problem over again, now with $L=10$. Note that choosing a larger value of L would require solving the optimization problem up to a (much) larger precision; as explained before, the value $L=10$ is quite sufficient for our purposes.

If the signs are to be computed as well, this would increase the dimension of the search space even more. However, because of their discrete nature, they are treated differently: First, a solution is computed corresponding to the common rescaling case. This solution is then improved by solving the directional-dependent rescaling case for unknown signs. As the result is pretty close to the final solution, the signs can be determined at this point. Finally, a more detailed calculation of the homogeneity lengths is carried out with the given signs, first with $L=\sqrt{10}$ and then with $L=10$. This procedure circumvents the need to solve a mixed continuous-discrete multi-dimensional optimization problem.

The optimization processes needed for iteratively establishing the homogeneity parameters are computationally expensive: The gradients at each point have to be computed a number of times. One way to accelerate this optimization would be to use the solution obtained in the previous point as

a starting guess for the optimization at the next point. This is advantageous mostly in cases where the gradients are computed with a rather high time resolution, since then the homogeneity lengths do not change much from point to point. For testing purposes, however, we have not followed this strategy here: By keeping the computation at each point independent from that in the other points it is easier to evaluate the correctness and the efficiency of the optimization processes.

5 Algorithm performance

In this section we illustrate the performance of LSGC-AS with the error estimation heuristics proposed above.

5.1 Gradient of a scalar field: a planar transition layer

In a first series of tests we consider a planar interface in a scalar field, e.g. plasma density (see Fig. 5). The interface has its normal direction along x and is characterized by a smooth transition profile of the form

$$n(x) = n_l \frac{1 + \mathrm{erf}(x/D)}{2} + n_r \frac{1 - \mathrm{erf}(x/D)}{2}.$$

The two asymptotic densities are chosen as $n_l=0.1\,\mathrm{cm}^{-3}$ and $n_r=10\,\mathrm{cm}^{-3}$ and the characteristic half-thickness is $D=100\,\mathrm{km}$. Synthetic data have been produced for multi-spacecraft constellations, in which the spacecraft fly through the interface along parallel straight lines from the left ($x<0$) to the right ($x>0$) with a constant speed $v_x=1\,\mathrm{km\,s}^{-1}$, while moving along the y-direction with $v_y=0.5\,\mathrm{km\,s}^{-1}$; there is no motion along z. The synthetic data sample the structure at 5 s resolution. Note that there is only a single direction of varying curvature in this example.

We first consider the case of a 4-spacecraft constellation that is about 150 km wide, on the order of the layer thickness. The spacecraft therefore cross the boundary with delays that are less than the crossing duration (Fig. 5a). The relative errors on the data have a normal distribution with a three-standard-deviation range of 5% (constant relative error). The x-, y-, z-, and t-axes are taken to be the homogeneity directions. The reference homogeneity lengths have been chosen isotropic in space, with $l_x^{(0)}=l_y^{(0)}=l_z^{(0)}=1000\,\mathrm{km}$, and the reference homogeneity time scale is $l_t^{(0)}=600\,\mathrm{s}$. The approximation error scaling factor is $f_c=1\,\mathrm{cm}^{-3}$. The gradients were computed with a selection limit $\sigma=1000$.

The least-squares gradient of (the logarithm of) the density has been computed every 30 s. Figure 6a–d shows the space-time gradient components obtained with different heuristic techniques for estimating the approximation error. Each of these techniques provides the gradient with a different total error estimate. The figure plots the results for global χ^2 optimization (orange), local χ^2 optimization (red), solution-adaptive common rescaling (green), direction-dependent rescaling with curvature while requiring
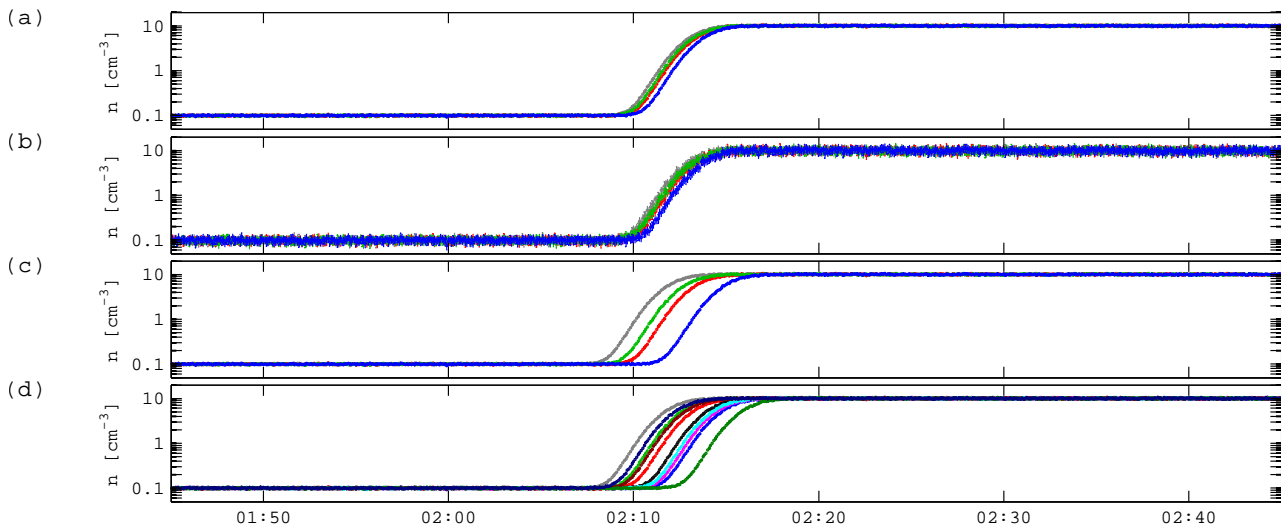
Fig. 5. Multiple spacecraft crossing a planar density interface with a half-width of 100 km. The density changes by two orders of magnitude. **(a)** 4 spacecraft, small separation (constellation scale ∼150 km) with relative measurement errors of 5%. **(b)** Idem, with 20% measurement errors. **(c)** 4 spacecraft, large separation (∼600 km), 5% errors. **(d)** 10 spacecraft, large separation (∼600 km), 5% measurement errors.

$\lambda_2 = \lambda_3$ (blue), and direction-dependent rescaling with curvature (black). The magenta lines correspond to the exact solution. All techniques recover the $\partial n/\partial x$ profile and find that the other gradient components are zero within the error margins. While the gradients obtained with these techniques are very similar, the error estimates are different. Obviously, the error bars are largest on the high-density side of the boundary, since the absolute density errors are two orders of magnitude larger there. The error bars on $\partial n/\partial x$ and $\partial n/\partial y$ are found to have the same order of magnitude, while those on $\partial n/\partial z$ are larger. This is because $\partial n/\partial z$ is obtained from differences over the spacecraft separation distance along z, while the x- and y-derivatives are obtained with a longer baseline as the spacecraft are moving along those directions.

Global χ^2 optimization is a robust and fast technique, just like local χ^2 optimization and solution-adaptive common rescaling. Figure 6e displays the relative homogeneity rescaling factor λ for these three methods. Because of our choice of $l^{(0)}$, λ expresses the homogeneity lengths in units of 1000 km as well as the homogeneity time as a multiple of 600 s. By construction, the global χ^2 technique provides a constant value of λ over the whole time interval, so it is not able to adapt to space-time variations in the way the local χ^2 heuristic and the common rescaling technique do: Those techniques correctly infer a minimum in the homogeneity scale as one crosses the density interface. The minimum $\lambda \approx 0.02$ corresponds to a spatial homogeneity domain diameter of 40 km, less than the layer half-thickness, as expected. In fact, the local techniques find two minima. The first one corresponds to the gradient change at the left edge of the transition. The second one is located at the center

of the transition, where the gradient reaches its peak value. There is no minimum at the right edge of the transition as this edge is smoother than the left edge when viewed in terms of the logarithm of the density. Note that the λ-value for global χ^2 optimization is basically set by the most difficult region encountered. The method therefore produces correct results, but fails to exploit the smoothness of the solution elsewhere, thereby overestimating the error margins dramatically. The local techniques, on the contrary, demonstrate the power of automatic approximation error estimation: As the homogeneity lengths adapt to the changing situation, the curvature errors (and hence the total error margins) are estimated in a more realistic fashion.

Relying on a common λ for all dimensions is adequate if the relative proportions of the reference homogeneity scales are appropriate for the whole time interval. If not, direction-dependent rescaling is advisable, although it is more time-consuming. Since the curvature signs can be obtained at nearly no additional cost, only direction-dependent rescaling with curvature is to be used in practice. Knowing the curvature signs allows a more precise estimation of the curvature error. Since in the present example the true situation has a nonzero curvature in a single dimension only (along x), both types of methods are essentially equivalent. We have tried both the heuristic that imposes $\lambda_2 = \lambda_3$ (corresponding to the y- and z-directions) to limit the number of parameters to be determined, and the full direction-dependent heuristic with four homogeneity scales. The two techniques give similar results, while the computing times are not very different. Both techniques find that λ_1 (Fig. 6f and h, orange), the scale along the x-direction, as being the smallest scale. As
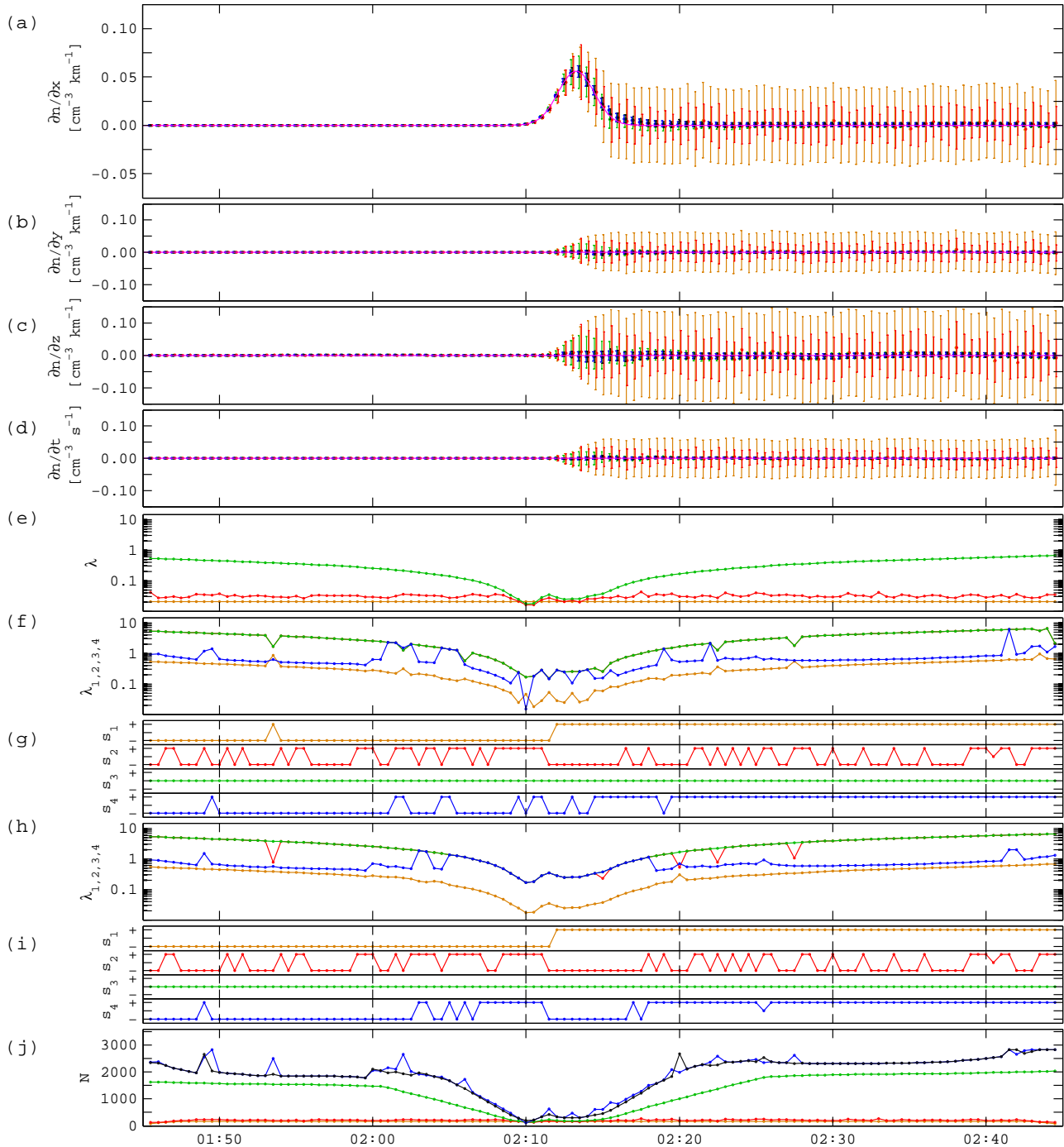
Fig. 6. Least-squares gradient at a density interface (4 spacecraft, small separation, 5% error). Panels **(a–d)** show the x-, y-, z-, and t-derivatives obtained with automatic error estimation using global and local χ^2 optimization (orange and red), common rescaling (green), and direction-dependent rescaling with curvature (with $\lambda_2 = \lambda_3$, blue, and the general case, black). Magenta lines indicate the exact gradient. All methods compute the gradients at the same instants; the data have been offset a little bit in time to improve visibility. **(e)** Relative homogeneity scale for global and local χ^2 optimization (orange and red), and for common rescaling (green). **(f)** and **(g)** Relative homogeneity scales and curvature signs for direction-dependent rescaling with curvature and $\lambda_2 = \lambda_3$ (orange: λ_1, along x; red: λ_2, y; green: λ_3, z; blue: λ_4, t). **(h)** and **(i)** Idem for the general direction-dependent case. **(j)** Number of equations (global and local χ^2 optimization: orange and red; common rescaling: green; direction-dependent rescaling with curvature: $\lambda_2 = \lambda_3$, blue, and the general case, black).

explained in Sect. 4.2, curvature estimation depends strongly on how well the homogeneity domain is covered by the observations. Because of the limited size of the space-time volume that is sampled by the four spacecraft, it is impossible to infer that λ_2, λ_3 and λ_2 are actually infinite, but one can set lower limits. By construction, the four curves in Fig. 6f are situated in a band with a range $\lambda_{min} \leq \lambda_k \leq 10\lambda_{min}$, where λ_{min} is the value obtained by the common rescaling heuristic. The lengths along y (red) and z (green) are at the upper limit of the band, while the relative time scale λ_4 (blue) roughly constant and always inside the band. The curvature signs s_k are shown in Fig. 6g and i. The signs s_2, s_3, and s_4 are ill-defined since the residuals in those directions are small, as indicated by the large homogeneity scales. The important sign here is s_1, the curvature in the x-direction: s_1 switches its sign at the center of the transition in $\log n$.

The computational complexity of the techniques is to a large extent determined by the size of the overdetermined systems that have to be solved to obtain the gradients. Figure 6j shows how the number of equations reflects the changes in the homogeneity scales.

Repeating these calculations with data onto which no measurement errors have been superposed, gives almost the same results (not shown). Larger measurement errors, however, may have an effect on the ability of the heuristics to detect the homogeneity scales properly. The above example has been reconsidered, now for the case of 20% measurement errors (see Fig. 5b). The results are summarized in Fig. 7. The correct $\partial n/\partial x$ profile is retrieved despite the larger measurement errors. The gradient components are fluctuating somewhat, but remain compatible with the exact solution given the larger error margins. Local χ^2 optimization systematically produces smaller homogeneity lengths (Fig. 7e, red curve) and correspondingly larger error margins on the gradients, with a certain degree of variability. The λ of common rescaling (Fig. 7e, green) and the λ_k and s_k of both direction-dependent rescaling variants (Fig. 7f–i) are largely the same as for the 5% measurement error case, be it with some random fluctuations: These heuristics are sufficiently robust so as not to be fooled too much by the large measurement errors.

Problems with random measurement errors can be partially avoided by first time-averaging the data on a time scale τ that is small enough not to blur the structures at hand:

$$\langle f(t_i) \rangle = \sum_{t_k \in \mathcal{W}_i} f(t_k)/n_i,$$

where n_i denotes the number of data points in the averaging window $\mathcal{W}_i = [t_i - \tau/2, t_i + \tau/2]$. This makes sense only if the sampling frequency if fairly high ($n_i \gg 1$). This reduces the errors on the time-averaged measurements:

$$\delta\langle f(t_i) \rangle^2 = \frac{1}{n_i^2} \sum_{t_k \in \mathcal{W}_i} \delta f_{m_k}^2 + \frac{1}{n_i(n_i-1)} \sum_{t_k \in \mathcal{W}_i} (f_k - \langle f(t_i) \rangle)^2,$$

which takes into account both the measurement errors (of diminishing importance as n_i grows) and the time-variability of

the observed quantity (assuming this variability to be gaussian, requiring an estimate of the standard deviation of the distribution). The data set subsequently has to be resampled at roughly the same time scale: Resampling it with a lower time resolution (larger time scale) would disregard some of the data, while data resampled at a higher time resolution (smaller time scale) or not resampled at all no longer are statistically independent. In summary: There are less data, but they are more precise. In the present example, we have carried out a modest smoothing of the 5 s data over a time scale $\tau = 30$ s, followed by a resampling at 15 s resolution. Figure 8a–d does not show a qualitative difference in the gradients, but especially for local χ^2 optimization the error margins are smaller. The heuristics work better so that the fluctuations in the values of the homogeneity scales and the curvature signs are suppressed somewhat (Fig. 8e–i). The resampling procedure reduces the number of available data to about one third of the original number as reflected in the overdetermined system size (Fig. 8j); the gradients are therefore computed significantly faster. Note that the resampling time scale should not be too large: Enough data points should be left to ensure a smooth behaviour of the target function $\mathcal{F}$ as a function of the homogeneity parameters (see the discussion on the danger of getting trapped in local minima in Sect. 4.4). In conclusion, a certain degree of a priori smoothing can efficiently remove a significant fraction of the measurement errors, thus facilitating error estimation and speeding up the computation.

We have repeated the calculations for a 4-spacecraft constellation with a larger spatial dimension of 600 km, and with measurement errors of 5% (Fig. 5c), again smoothed at a 30 s time scale and resampled at 15 s resolution. The spacecraft separation is now larger than the layer thickness. The computed gradients (Fig. 9a–d) deviate somewhat from the exact solution: Spurious y- and z-components appear in the layer, a phenomenon that is well known for the standard CGC technique. The techniques indicate quite precise results outside the layer, where the length scales are large anyhow and the larger separation allows a more precise evaluation of the gradient. Inside the layer, the computed gradients carry larger error bars as expected. In fact, the error bars that are associated with global and local χ^2 optimization and with common rescaling are larger than the gradient, so that one would deem the observed result to be consistent with zero gradient. Only the direction-dependent techniques, because of the more precise underlying approximation error model, are able to produce results with smaller error estimates that clearly establish this gradient as being significant. The homogeneity scales and curvature signs are found to be similar as before (Fig. 9e–i), confirming that the heuristics are fairly robust, although at the center of the layer λ_1 is not always the smallest. The same minimum length scale is found as before, and the sign change for s_1 is again correctly established.

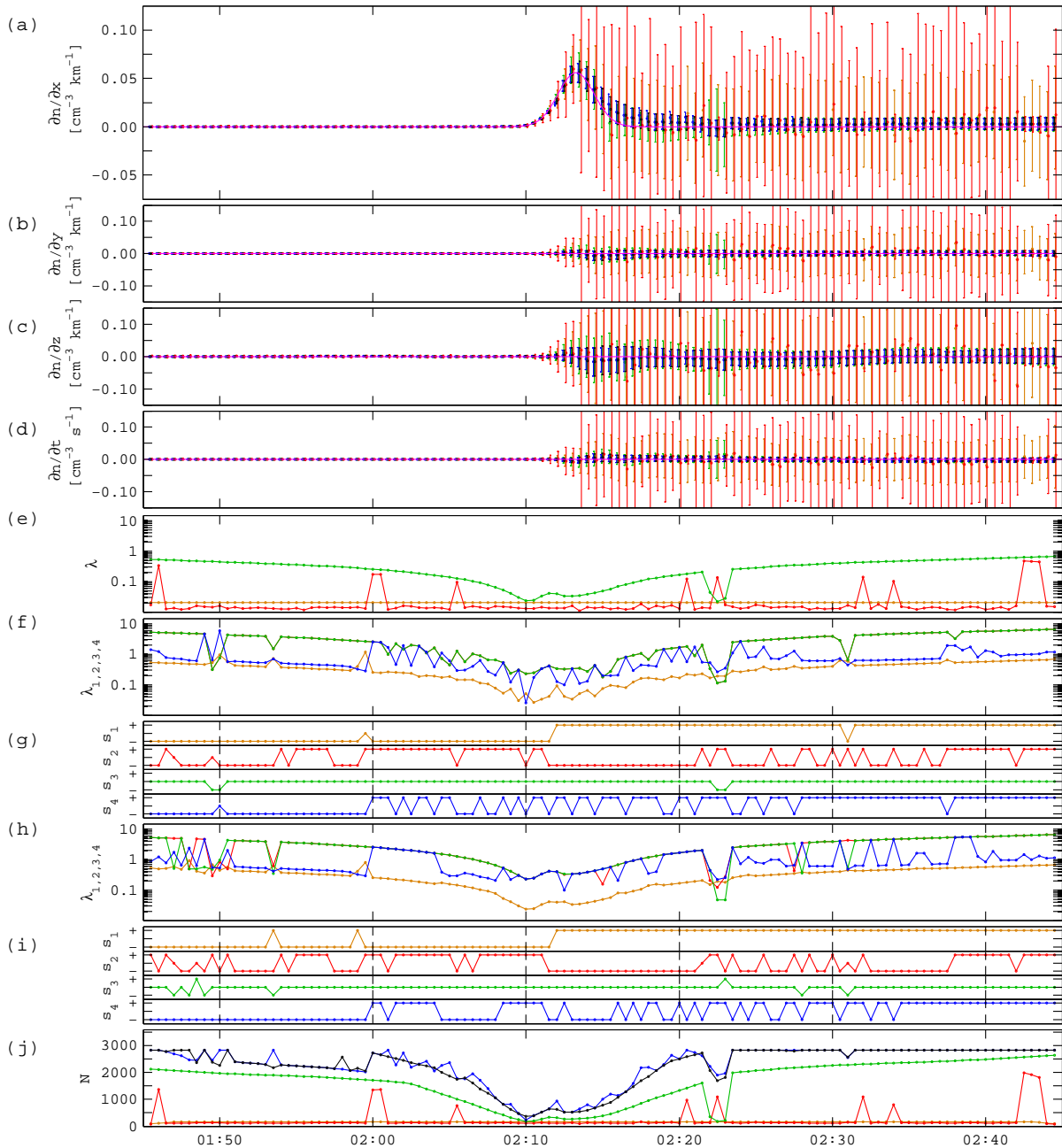It is particularly interesting to investigate how the heuristics behave for constellations with more spacecraft, for

Fig. 7. Least-squares gradient at a density interface (4 spacecraft, small separation, 20% measurement error). See caption of Fig. 6. To facilitate comparison, the same axis scales have been used.

instance for $K=10$ spacecraft. The profiles for such a case are shown in Fig. 5d. More spacecraft lead to a larger set of data with a better sampling of the homogeneity domain. The 10-spacecraft configuration considered here is an extension of the 600 km-separation 4-spacecraft configuration of the previous example; spacecraft have been added between and around the original set, which should help to better define both the small and the large homogeneity scales.

The data are smoothed and resampled as before. The gradient components are found to be determined quite accurately (Fig. 10a–d), with reasonable error bars. The direction-dependent methods recover the length scales and curvature signs pretty well; there are still some problems at the center of the layer (Fig. 10e–i) where λ_1 is not always identified as the smallest scale. The larger number of data points involved in these calculations is reflected in Fig. 10j.
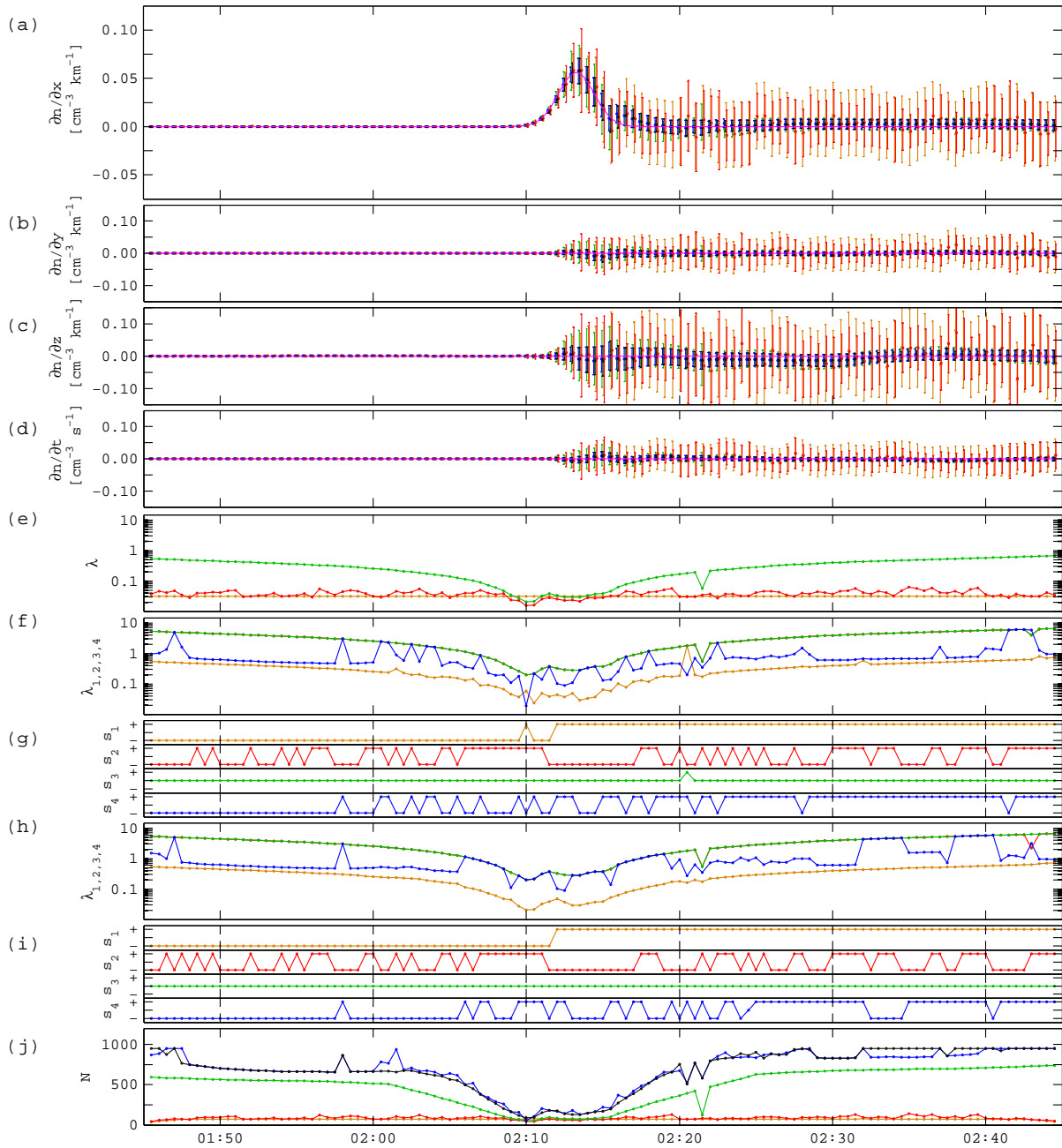
Fig. 8. Least-squares gradient at a density interface (4 spacecraft, small separation, 20% measurement error, smoothed at a 30 s time scale, resampled at 15 s resolution). See caption of Fig. 6.

When one attempts to compute gradients with spacecraft that are too far apart, LSGC-AS finds that there are not enough measurement points in the homogeneity domain: The data points closest to x_0 already carry large approximation errors. The resulting gradients therefore have excessive error bars. In our implementation, such results are simply discarded.

5.2 Gradients of a vector field: a dipole with a ring current

The application of LSGC to vector fields has been demonstrated by De Keyser et al. (2007). They have shown that for divergence-free vector fields, such as the magnetic field, the difference between solutions obtained with or without imposing the divergence-free constraint is not very large. That should not be a surprise, as imposing this constraint reduces
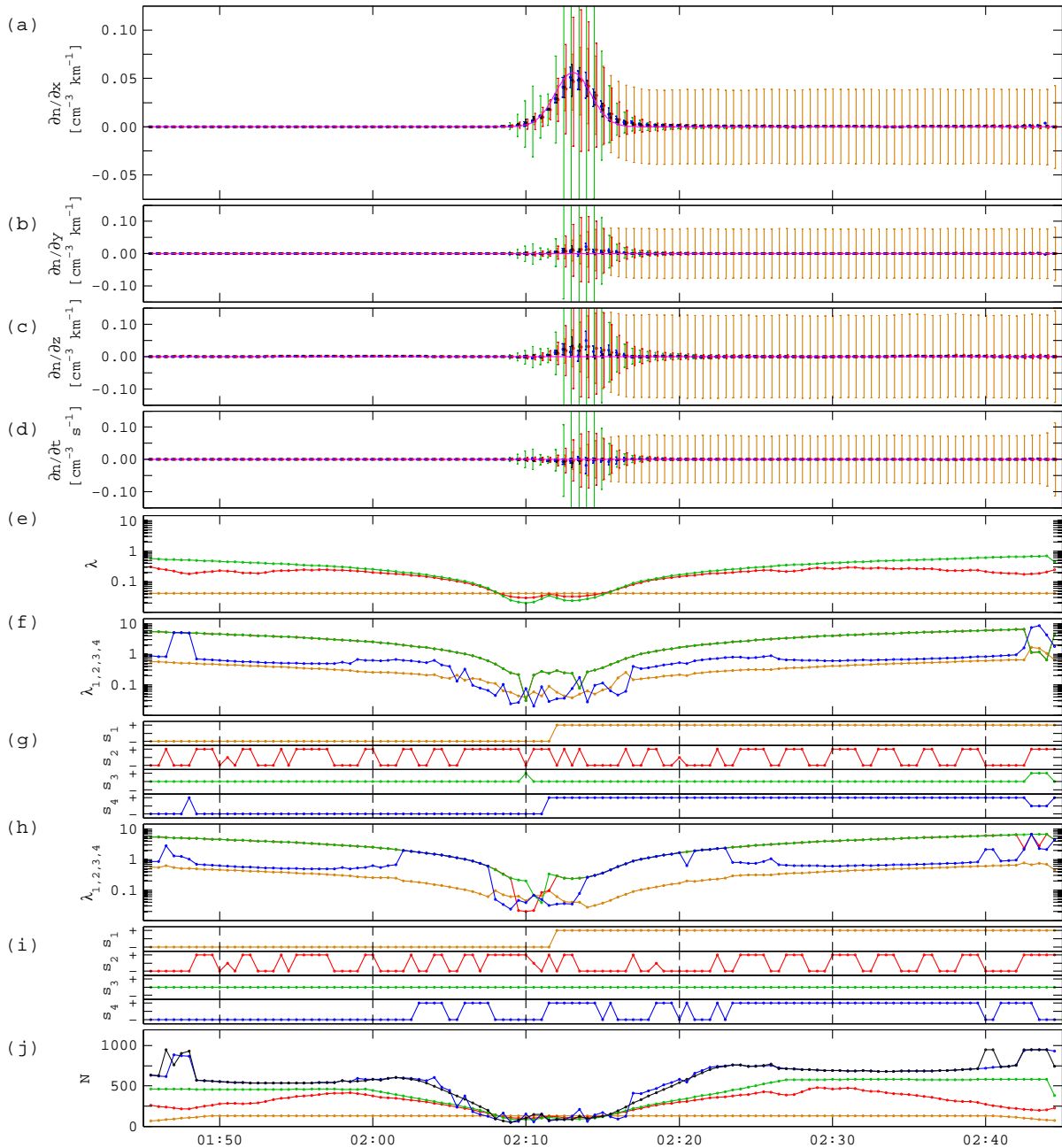
Fig. 9. Least-squares gradient at a density interface (4 spacecraft, medium separation, 5% measurement error, smoothed at a 30 s time scale, resampled at 15 s resolution). See caption of Fig. 6.

the dimension of the solution space from $M=15$ to $M=14$ only. Nevertheless, adding the $\nabla \cdot \boldsymbol{B}=0$ condition does improve the realism of the result. Its usefulness might be more pronounced when the number of measurements is limited. Moreover, including such a constraint in LSGC is not difficult. Therefore only this improved curlometer is used here.

Given its geophysical importance, we examine how this constrained least-squares curlometer would benefit from ap-

proximation error estimation. Consider a dipole magnetic field with its axis along z, on top of which the field induced by a toroidal ring current is superposed. The ring current density is defined by

$$j(r, z) = \frac{j_0}{1 + \left[\left(\frac{r-r_0}{a} \right)^2 + \left(\frac{z}{b} \right)^2 \right]^4},$$
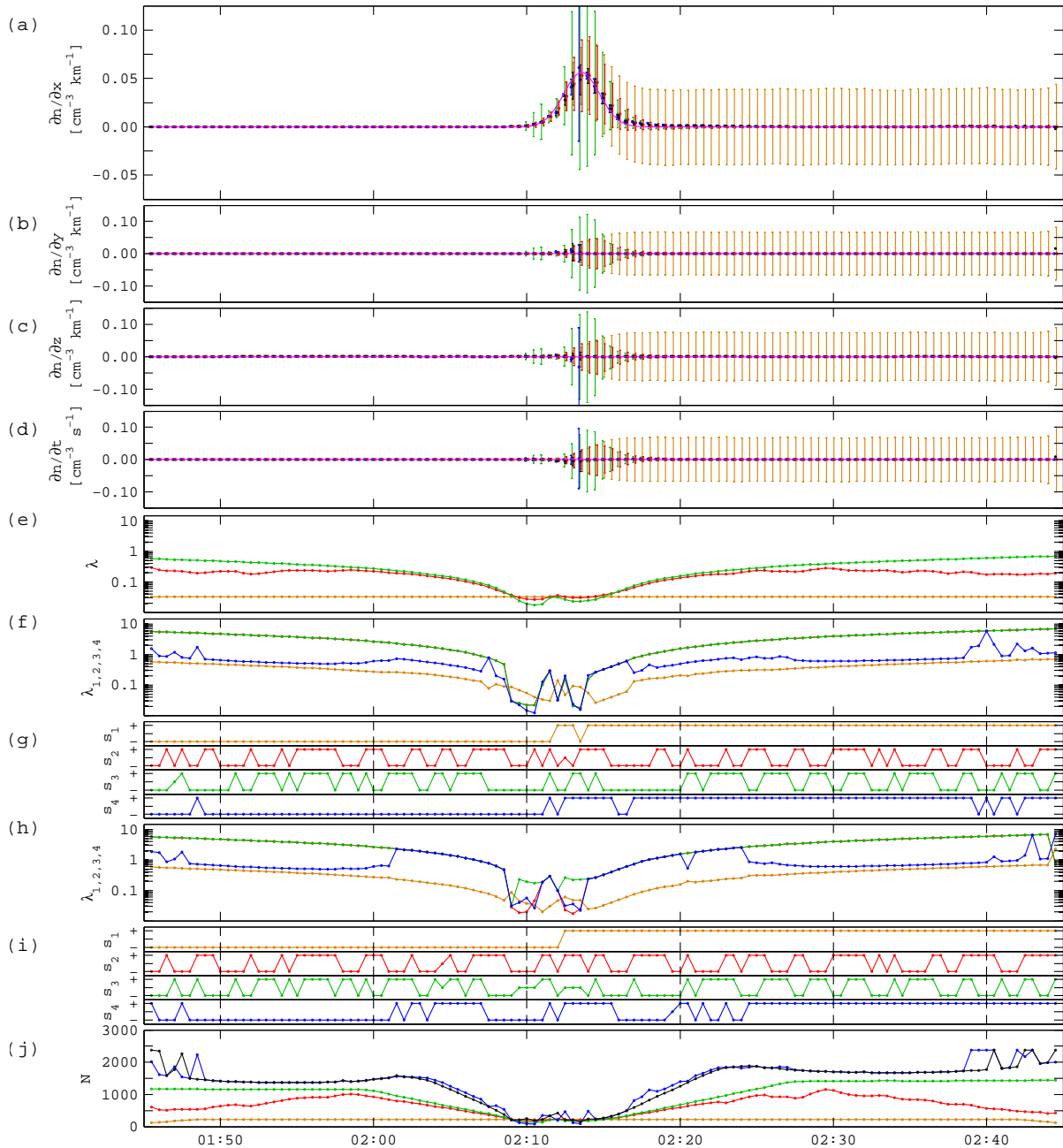
Fig. 10. Least-squares gradient at a density interface (10 spacecraft, medium separation, 5% measurement error, smoothed at a 30 s time scale, resampled at 15 s resolution). See caption of Fig. 6.

where the peak current density $j_0 = 20 \, \mathrm{nA \, m^{-2}}$ is centered at $r_0 = 5 \, R_E$, with the current distribution having a broad flat maximum that extends over $2a = 2 \, R_E$ radially and $2b = 6 \, R_E$ in the axial direction, while the current rapidly falls off outside that region. This toroidal current distribution and the associated axisymmetric field are illustrated in Fig. 11. Such a ring current distribution can be regarded as a set of infinitely narrow current loops. An analytical expression is known for

the induced magnetic field of such a current loop in terms of the complete elliptic integrals of the first and second kind. Integrating over the current distribution gives the total induced field. Figure 12 shows the magnetic field profiles that would be recorded by four spacecraft flying at 150 km separation through this dipole with ring current. The spacecraft to follow straight parallel trajectories, moving from outside this magnetosphere towards perigee at the end of the interval
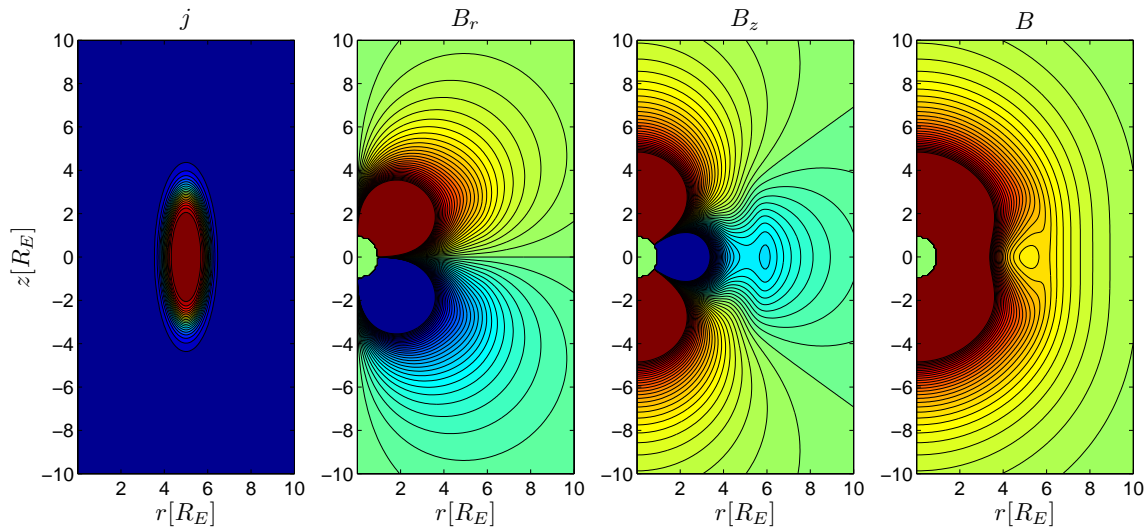
Fig. 11. Axisymmetric ring current distribution j, with a peak current of $20\,\mathrm{nA\,m^{-2}}$, and the magnetic field it induces, superposed onto a (non-tilted) terrestrial dipole field. From left to right: current density j, radial field B_r, axial field B_z, and field strength B. Magnetic field contours are drawn at 20 nT intervals; no contours are drawn close to Earth as the magnetic field rises sharply there.

with a velocity $[2, 0.5, 0]\,\mathrm{km\,s^{-1}}$ at $1\,R_E$ above the equatorial plane. These trajectories sample the weak magnetic fields far from the dipole axis, cross the ring current, and finally observe the strong fields in the inner magnetosphere. The insets show the data for a small time interval when the spacecraft enter the ring current region. The spacecraft, as they are closely spaced, observe fields that differ by 1 nT or less; as the value of j_0 reflects a typical ring current intensity, this is quite realistic. Nevertheless, the high measurement precision of 0.1 nT should make gradient computation feasible. The goal of our numerical experiment is to verify whether such current densities can indeed be recovered from such magnetic field signatures. Being able to measure such ring current densities is of considerable importance (Vallat et al., 2005).

Because of the difficulty of the problem, we have first smoothed the data at a 60 s time scale and resampled them at 30 s resolution. Figure 13 displays the results obtained with the divergence-free vector field LSGC-AS techniques, with gradients computed every 3 min. The three components of $\nabla \times \boldsymbol{B}$ in Fig. 13a–c have been obtained with LSGC in which the gradients of B_x, B_y, and B_z are computed simultaneously, coupled through the condition $\nabla \cdot \boldsymbol{B} = 0$. The homogeneity directions have been chosen to align with the magnetic field, which is a reasonable thing to do inside the magnetosphere, although the ring current in the example is not strictly aligned with the magnetic field. The figure shows the results for the case of common rescaling (green curves) and for direction-dependent rescaling with curvature and with four different scales corresponding to the homogeneity directions (λ_1 and λ_2 in the plane perpendicular to $\boldsymbol{B}$, λ_3 along $\boldsymbol{B}$, and

λ_4 for the time dimension). For comparison we also plot the exact value of $\nabla \times \boldsymbol{B}$, which is proportional to the current density as $\boldsymbol{j} = \nabla \times \boldsymbol{B}/\mu_0$ in a steady field. The computed x- and y-components of the curl clearly trace the correct profiles as the ring current region is traversed, with a gradual rise of the current from zero to a broad plateau and back to zero again; the z-component remains zero. The error margins for direction-dependent rescaling are necessarily smaller than those of common rescaling (in which the minimum scale is used in all directions). In the ring current region, the error margins are sufficiently small so that the amplitude in the nonzero-current region is detected as being statistically significant. As the spacecraft approach the Earth, the error margins grow, since the second derivatives of the field components are much larger there. The exit of the spacecraft from the ring current region is therefore only marginally significant. It should be stressed here that the 3-standard-deviation error margins indeed comprise the deviations between computed and true values of the curl everywhere. Figure 13d–i show the homogeneity scales and curvature signs determined by LSGC-AS for the gradients of B_x, B_y, and B_z. It is not surprising that the scales remain in a band (corresponding to factor $L = 10$) that progressively decreases as the spacecraft approach Earth. Since the same reference scales have been used as in Sect. 5.1, the smallest curvature scale is seen to vary from roughly 1000 km ($\lambda \approx 1$) down to 100 km ($\lambda \approx 0.1$) as the field variations become stronger near perigee. The number of equations in the coupled overdetermined system is given in Fig. 13j: More than 5000 equations (corresponding to 5000/3 data points) are involved in the computations inside the ring current layer. Because of the data compression
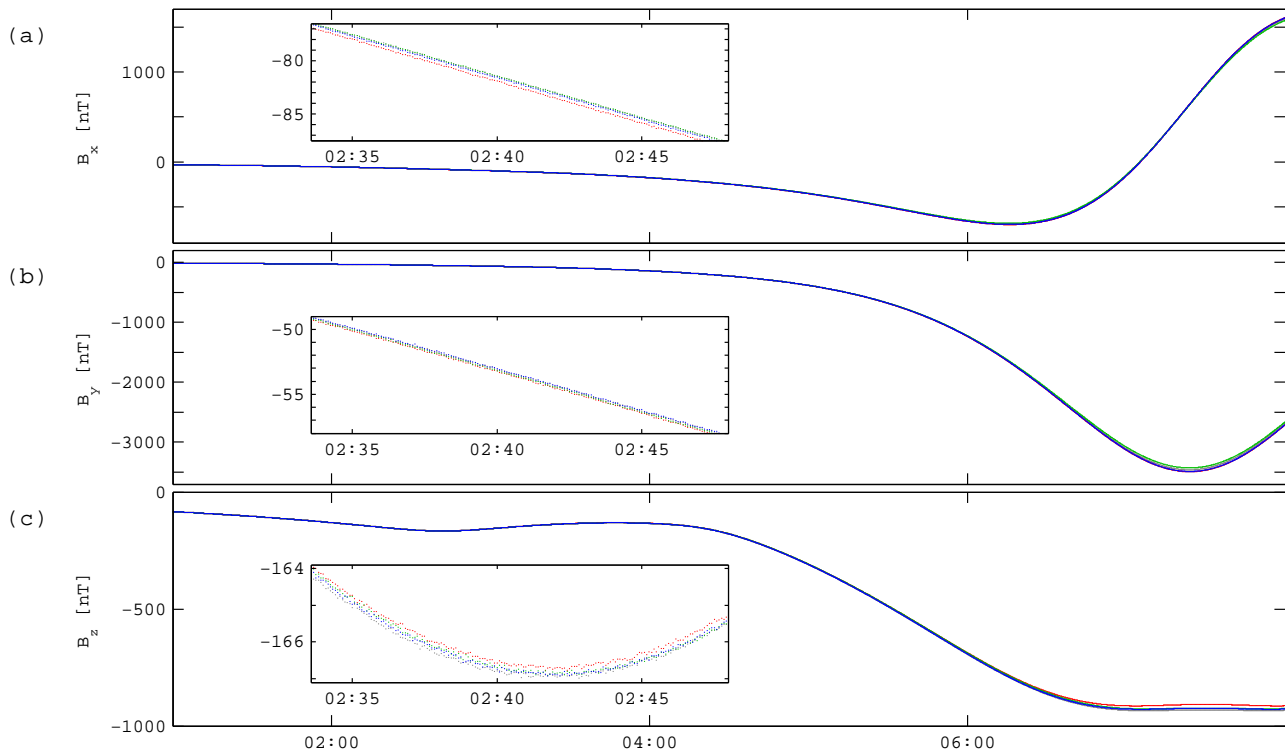
Fig. 12. Magnetic field profiles for a 4-spacecraft configuration at 150 km separation flying through the dipole with ring current depicted in Fig. 11. The trajectories are parallel straight lines, scanning along x and y with speeds of 2 and 0.5 km s^{-1}, respectively, at a constant z of 1 R_E. The figure shows the field as the spacecraft move in towards perigee. The insets zoom in on the magnetic field data around the time when the spacecraft enter the ring current region; the magnetic field differences are of the order of 1 nT or less while the fields are supposed to be measured with a 0.1 nT precision.

by a factor 6 arising from the smoothing/resampling preprocessing step, each overdetermined system there combines information from more than 10 000 data points.

6 Conclusions

Least-squares methods for computing the gradient from multi-spacecraft data offer a number of advantages over the classical 4-point gradient. In particular, LSGC uses information from a large number of data points to obtain a more precise result than CGC. It also provides a total error estimate on the computed gradient. All of this, however, depends on the availability of proper estimates for the approximation error due to the non-constant nature of the gradient over the set of points that is used to compute the gradient from. In this paper, we have proposed and tested a number of heuristics to automatically infer such approximation error estimates, leading to various LSGC-AS techniques with adaptive estimation of the homogeneity scales. These heuristics are built from simple models of the approximation error that are expressed in terms of the homogeneity parameters. After computing the gradient starting from an initial choice of parameter values,

the residuals are used to improve these homogeneity parameters. An optimization problem can be formulated where the optimum corresponds to the most appropriate choice of the homogeneity parameters. The result is a gradient that is computed from the appropriate set of data points. The associated total error estimates are usually reliable. Even in those cases where some of the homogeneity lengths are too small to be properly sampled by the spacecraft configuration and the error estimates cannot be determined with much precision, the error margins are large so that the heuristics at least indicating where the gradient should be considered with more caution. An obvious advantage of LSGC-AS is that the user does not have to provide any input on the homogeneity scales.

As has been shown, the precision of the computed gradient is determined largely by the measurement errors at the sampling points that are being used, but the construction of that set is determined by the approximation errors (the homogeneity parameters). The quality of the solution therefore depends on the measurement errors as well as on the distribution of points over the homogeneity domain, which is determined by the number of spacecraft, their separations, their velocity, and the sampling rate. While LSGC in general, and
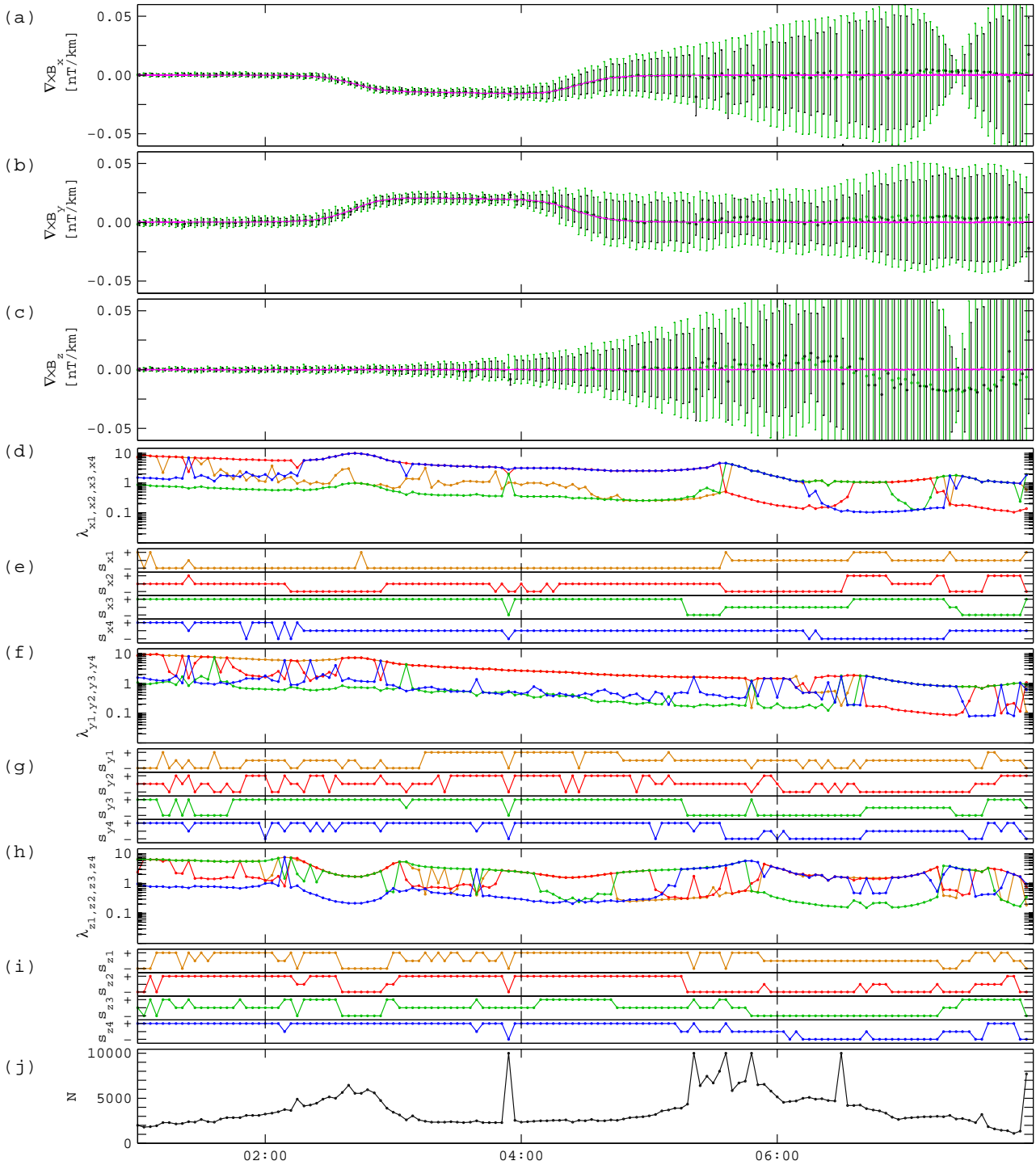
Fig. 13. Curl of the magnetic field, $\nabla \times \boldsymbol{B}$, obtained from least-squares gradient computation for the field components, subject to the condition $\nabla \cdot \boldsymbol{B} = 0$, in the dipole with ring current illustrated in Figs. 11 and 12 for a 4-spacecraft configuration. The data were smoothed at a 60 s time scale and resampled at 30 s resolution. **(a–c)** LSGC calculation of the x-, y-, and z-components of $\nabla \times \boldsymbol{B}$ with common rescaling (green) and direction-dependent rescaling (black), compared to the exact solution (magenta). **(d)** and **(e)** Relative homogeneity lengths and curvature signs for direction-dependent rescaling for ∇B_x (λ_1 orange, λ_2 red, λ_3 green, λ_4 blue). **(f)** and **(g)** Idem for ∇B_y. **(h)** and **(i)** Idem for ∇B_z. **(j)** Number of equations in the coupled overdetermined system for the three gradients.

LSGC-AS in particular, are independent of the actual number of spacecraft, the results that one obtains in any particular situation obviously do depend on the spacecraft configuration.

The proposed heuristics seem to work well if the measurement errors are small. They also work reasonably well if the measurement errors are larger (except for the χ^2-based techniques), but in such cases it can be recommended to perform some a priori time-averaging on a time scale that is large enough to suppress the random measurement errors significantly, but small enough so that it does not wash away details at the homogeneity scales. Correspondingly resampling the data at a coarser time scale is a good idea as it can speed up the computations considerably. The simpler error estimation heuristics are robust, but tend to associate rather large error margins with the computed gradients. The more sophisticated techniques produce gradients that are more precise, with narrower error margins. While the approach in this paper has been to consider the error margins to be at the 3-standard-deviation level, the computed gradients can sometimes deviate a little more from the true solution. That can be the consequence of the simplicity of the error models, of ignoring the cross-correlations between the approximation errors, and of ignoring the covariances between the solution components. In summary: LSGC-AS produces 3-standard-deviation errors, but because of various simplifications the actual errors occasionaly can be a little larger. Repeating the computations with different heuristics and comparing their respective results could help in identifying such occasional problems.

A strong point of LSGC-AS is that it is straightforward to include constraints. (1) Physical constraints: The most notable example of a physical constraint is the introduction of the divergence-free condition for vector fields such as the magnetic field, leading to the improved curlometer proposed by De Keyser et al. (2007) and combined here with the error estimation heuristics. (2) Geometric constraints: Such constraints express that the gradient is perpendicular to a given direction. Each geometric constraint fixes a homogeneity direction and amounts to taking the corresponding homogeneity length to be infinitely large (De Keyser et al., 2007, Appendix A1). This simplifies the error estimation heuristics as the set of parameters becomes smaller. A first example is requiring the gradient to be perpendicular to the local magnetic field. Such a condition reduces the dimensionality of the problem, and thus the number of spatial unknowns and also the number of parameters in the error estimation optimization process. A second example is to assume (approximate) time stationarity of the observed structures by considering a frame moving with the field, requiring the total derivative of the observed field to be (approximately) zero. This implies a choice of homogeneity directions that involves the convection speed of the structures (De Keyser et al., 2007, Appendix A2); this speed might be known, for instance, from plasma measurements. (3) In a third type of constraints one specifies a dimension in which one has to be very selective.

One then identifies a homogeneity direction in which the homogeneity scale is vanishingly small. For example, if rapid temporal changes are possible, only simultaneous measurements should be used for computing the gradient; this can be enforced by setting $l_t^{(0)} \to 0$. The optimization process automatically finds that rescaling such a short time scale by a finite factor does not change the results: For four spacecraft, and assuming the other homogeneity scales to be infinitely large, this reduces to the standard CGC method (Harvey, 1998; Chanteur, 1998; Chanteur and Harvey, 1998).

The choice of the homogeneity directions should be tailored to the problem at hand. This is true in general, but especially as soon as some of the above-mentioned constraints are involved. Once such a choice has been made, the heuristics that have been explored here are able to automatically provide the corresponding homogeneity scales. If an ill-advised choice of homogeneity directions has been made, that does not produce erroneous results: It only means that, with a better choice, more precise results and narrower error margins could have been obtained.

In summary, least-squares gradient computation with automatic homogeneity scale estimation (LSGC-AS) and the corresponding curlometer are essential multi-spacecraft data analysis tools: Up to now, LSGC-AS is the only way to obtain at the same time the gradient value and the corresponding uncertainty. Since analyzing gradients of magnetospheric fields with limited spacecraft constellations is synonymous to working at the limits of precision, it is of utmost practical importance to have a proper estimate of the error. In addition, LSGC-AS provides some insight as to where the errors come from by identifying the homogeneity scales as they vary throughout the magnetospheric regions that are traversed.

References

Balogh, A., Dunlop, M. W., Cowley, S. W. H., Southwood, D. J., Thomlinson, J. G., Glassmeier, K.-H., Musmann, G., Lühr, H., Buchert, S., Acuña, M. H., Fairfield, D. H., Slavin, J. A., Riedler, W., Schwingenschuh, K., Kivelson, M. G., and the Cluster magnetometer team: The Cluster Magnetic Field Investigation, Space Sci. Rev., 79, 65–91, 1997.

Balogh, A., Carr, C. M., Acuña, M. H., Dunlop, M. W., Beek, T. J., Brown, P., Fornaçon, K.-H., Georgescu, E., Glassmeier, K.-H., Harris, J., Musmann, G., Oddy, T., and Schwingenschuh, K.: The Cluster Magnetic Field Investigation: overview of in-flight performance and initial results, Ann. Geophys., 19, 1207–1217, 2001, http://www.ann-geophys.net/19/1207/2001/.

Chanteur, G.: Spatial Interpolation for Four Spacecraft: Theory, in: Analysis Methods for Multi-Spacecraft Data, edited by: Paschmann, G. and Daly, P. W., pp. 349–369, ISSI Scientific Report SR-001, 1998.

Chanteur, G. and Harvey, C. C.: Spatial Interpolation for Four Spacecraft: Application to Magnetic Gradients, in: Analysis Methods for Multi-Spacecraft Data, edited by: Paschmann, G. and Daly, P. W., pp. 371–393, ISSI Scientific Report SR-001, 1998.

Darrouzet, F., De Keyser, J., Décréau, P. M. E., Lemaire, J. F., and Dunlop, M. W.: Spatial gradients in the plasmasphere from Cluster, Geophys. Res. Lett., 33, L08105, doi:10.1029/2006GL025727, 2006.

De Keyser, J., Darrouzet, F., Dunlop, M. W., and Décréau, P. M. E.: Least-squares gradient calculation from multi-point observations of scalar and vector fields: Methodology and applications with Cluster in the plasmasphere, Ann. Geophys., 25, 971–987, 2007, http://www.ann-geophys.net/25/971/2007/.

Décréau, P. M. E., Fergeau, P., Krasnoselskikh, V., Lévêque, M., Martin, P., Randriamboarison, O., Séné, F. X., Trotignon, J. G., Canu, P., Mögensen, P. B., and Whisper investigators: WHISPER, A Resonance Sounder and Wave Analyser: Performances and Perspectives for the Cluster Mission, Space Sci. Rev., 79, 157–193, 1997.

Décréau, P. M. E., Fergeau, P., Krasnoselskikh, V., Le Guirriec, E., Lévêque, M., Martin, P., Randriamboarison, O., Rauch, J. L., Séné, F. X., Séran, H. C., Trotignon, J. G., Canu, P., Cornilleau, N., de Féraudy, H., Alleyne, H., Yearby, K., Mögensen, P. B., Gustafsson, G., André, M., Gurnett, D. A., Darrouzet, F., Lemaire, J., Harvey, C. C., Travnicek, P., and Whisper experimenters: Early results from the Whisper instrument on Cluster: an overview, Ann. Geophys., 19, 1241–1258, 2001, http://www.ann-geophys.net/19/1241/2001/.

Dunlop, M. W. and Balogh, A.: Magnetopause current as seen by Cluster, Ann. Geophys., 23, 901–907, 2005, http://www.ann-geophys.net/23/901/2005/.

Dunlop, M. W., Balogh, A., Glassmeier, K.-H., and Robert, P.: Four-point Cluster application of magnetic field analysis tools: The Curlometer, J. Geophys. Res., 107, 1384, doi:10.1029/2001JA0050088, 2001.

Dunlop, M. W., Balogh, A., Shi, Q.-Q., Pu, Z., Vallat, C., Robert, P., Haaland, S., Shen, C., Davies, J. A., Glassmeier, K.-H., Cargill, P., Darrouzet, F., and Roux, A.: The Curlometer and other gradient measurements with Cluster, Proceedings of the Cluster and Double Star Symposium, 5th Anniversary of Cluster in Space, ESA SP-598, 2006.

Harvey, C. C.: Spatial Gradients and the Volumetric Tensor, in: Analysis Methods for Multi-Spacecraft Data, edited by: Paschmann, G. and Daly, P. W., pp. 307–322, ISSI Scientific Report SR-001, 1998.

Robert, P., Dunlop, M. W., Roux, A., and Chanteur, G.: Accuracy of Current Density Determination, in: Analysis Methods for Multi-Spacecraft Data, edited by Paschmann, G. and Daly, P. W., pp. 395–418, ISSI Scientific Report SR-001, 1998a.

Robert, P., Roux, A., Harvey, C. C., Dunlop, M. W., Daly, P. W., and Glassmeier, K.-H.: Tetrahedron Geometric Factors, in: Analysis Methods for Multi-Spacecraft Data, edited by: Paschmann, G. and Daly, P. W., pp. 323–348, ISSI Scientific Report SR-001, 1998b.

Trotignon, J. G., Décréau, P. M. E., Rauch, J. L., Le Guirriec, E., Canu, P., and Darrouzet, F.: The Whisper Relaxation Sounder Onboard Cluster: A Powerful Tool for Space Plasma Diagnosis around the Earth, Cosmic Research, 41, 369–372, 2003.

Vallat, C., Dandouras, I., Dunlop, M., Balogh, A., Lucek, E., Parks, G. K., Wilber, M., Roelof, E. C., Chanteur, G., and Rème, H.: First current density measurements in the ring current region using simultaneous multi-spacecraft CLUSTER-FGM data, Ann. Geophys., 23, 1849–1865, 2005, http://www.ann-geophys.net/23/1849/2005/.