Atmos. Chem. Phys. Discuss., 9, C10551–C10558, 2010
www.atmos-chem-phys-discuss.net/9/C10551/2010/

Atmospheric Chemistry and Physics Discussions

ACPD

9, C10551–C10558, 2010

Interactive Comment

Interactive comment on "Technical Note: Variance-covariance matrix and averaging kernels for the Levenberg-Marquardt solution of the retrieval of atmospheric vertical profiles" *by* S. Ceccherini and M. Ridolfi

S. Ceccherini and M. Ridolfi

s.ceccherini@ifac.cnr.it

Received and published: 5 February 2010

We thank Dr. von Clarmann for his short comment (SC) that will permit to clarify important aspects that are not fully explained in the discussion paper.

Dr. von Clarmann claims that it is possible to demonstrate that converged Levenberg-Marquardt (LM) retrievals are characterized by the same covariance matrices and averaging kernels as Gauss-Newton (GN) retrievals. The proof is made using two approaches: the first one is based on Eq. (2) of the SC, the second is based on Eqs. (3-6).

In the final part of the SC the results obtained in the discussion paper are interpreted supposing that our retrievals are "non-converged".

In the following we discuss in detail the underlying hypotheses of the proof presented in the SC, and demonstrate that the results of the discussion paper are more general. Furthermore we provide evidence that our retrievals are well converged.

1 Analysis of the proof

Eq. (2) of the SC (first approach) is a vector equation and consists of n scalar equations with n unknowns. Therefore, it is possible that this vector equation uniquely determines the n elements of $\mathbf{x}$ (as Dr. von Clarmann states), however this is not always the case. Indeed, if the problem is ill-posed or ill-conditioned, the n equations are not independent from each other and they are not sufficient to determine an unique solution. In these cases a whole class of solutions is compatible with Eq. (2) and the iterative procedure described by Eq. (1) identifies one of these solutions. The achieved solution depends on the path followed by the minimization procedure in the parameter space and on the LM terms $\lambda_i \mathbf{D}_i$. If the inversion is ill-conditioned or ill-posed, the GN procedure is not able to minimize the cost function because the inversion of matrix $\left(\mathbf{K}_i^T \mathbf{S}_{\mathrm{y}}^{-1} \mathbf{K}_i + \mathbf{R}\right)$ either is not possible at all (ill-posed case), or the result significantly amplifies the existing errors (ill-conditioned case).

An excellent illustration of the inability of the GN method to find a solution in ill-conditioned cases can be obtained by implementing into a computer program the following example suggested at page 31 of Twomey (1977). Let n be equal to 2 and the observations vector be $(y_1, y_2)^T = (2, 4)^T$ with uncorrelated errors equal to 0.1. Assume the following forward model:

$$f_1 = x_1 + x_2 \tag{1}$$

ACPD
9, C10551–C10558, 2010

$$f_2 \quad = \quad 2x_1 + (2+\epsilon)x_2,$$

and try to find the solution $(\hat{x}_1, \hat{x}_2)^T$ using the iterative GN method starting from $(x_1, x_2)^T = (-10, 10)^T$. The minimum of the cost function (χ^2) is for $(x_1, x_2)^T = (2, 0)^T$, however, depending on the adopted numerical precision, there exists a value of ϵ below which GN does not converge. We tried the implementation of this exercise and found that in our computer setup (double precision variables and ifort compiler with standard flags for a 64-bit Xeon processor) the GN method fails to converge already for $\epsilon \leq 10^{-8}$. Of course, since the lack of convergence is caused by the numerical errors involved in the inversion of an ill-conditioned matrix, the value of ϵ below which convergence problems start depends on the numerical precision used.

Therefore, the retrieved values may depend on the path followed by the minimization process and there are cases in which the GN path does not reach convergence, which is instead reached by the LM path.

Now we consider the algebraic proof given by Eqs. (3-6) of the SC (second approach). The last step of Eq. (4) implies (again) the inversion of the matrix $\left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R}\right)$. This operation involves significant errors in case of ill-conditioning and is not possible at all if the problem is ill-posed. In these cases the equation:

$$\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} = \left(\mathbf{R} + \mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i\right) \mathbf{T}_{i+1} \qquad (2)$$

does not uniquely determine $\mathbf{T}_{i+1}$, but a whole class of matrices $\mathbf{T}_{i+1}$ can fulfill this equation. The matrix determined by the method described in the discussion paper determines a solution that depends on the path followed by the minimization procedure in the parameter space and on the LM terms $\lambda_i \mathbf{D}_i$. In ill-posed and ill-conditioned inversions the LM technique acts as an external constraint selecting one solution among all the possible ones. Therefore, as any other external constraint, the LM technique impacts both the covariance matrix and the averaging kernels of the solution. The

ACPD
9, C10551–C10558, 2010

formulas given in the discussion paper are correct in all cases: well-conditioned, ill-conditioned and ill-posed problems. Conversely, formulas (5) and (6) of the SC of Dr. von Clarmann are limited to the case of well-conditioned problems. In the case of well-conditioned problems the formulas of the discussion paper produce the same results as those of Dr. von Clarmann.

The differences that exist between our results and the expectations of Dr. von Clarmann can be explained by the fact that the test retrievals presented in the discussion paper do not use external constraints ($\mathbf{R} = \mathbf{0}$) and the pure GN inversion (i.e. without the LM term) is not sufficiently well-conditioned. This means that ill-conditioning amplifies both the numerical errors occurring in matrix inversion and the small errors in the Jacobian $\mathbf{K}_i$ (due to code optimizations) to a level that prevents the GN method from converging. This is the main reason why the proofs presented in the SC do not apply to our retrievals. We will modify the revised paper to make clear this concept.

2 Convergence

In the last paragraph (p.C9662) of the SC, Dr. von Clarmann argues that our LM retrievals could be non-converged. The non-convergence could be due to a large LM damping that produces a small chi-square variation far from convergence, hence erroneously triggering a successful convergence check.

First of all we would like to distinguish between “exact numerical convergence” and “physical convergence” (for this discussion we refer to Sect. 5.6.3 of Rodgers, 2000). Exact numerical convergence occurs when in subsequent iterations, the cost function to be minimized and / or the retrieval vector, do not change at machine precision. Exact numerical convergence is clearly expensive and unnecessary to achieve. What is physically meaningful, is to require that the difference between the solution and the

ACPD

9, C10551–C10558, 2010

Interactive Comment

state corresponding to the minimum of the cost function should be much smaller (e.g. one order of magnitude) than the retrieval error due to measurement noise. We call this latter condition “physical convergence”. According to Sect. 5.6.3 of Rodgers (2000), an appropriate condition that should be checked to detect whether the physical convergence has been achieved is:

$$\frac{\chi^2(\mathbf{x}_{i+1}) - \chi^2(\mathbf{x}_i)}{\nu} << 1. \tag{3}$$

with ν equal to the number of degrees of freedom of the χ^2 distribution. In our retrievals we stop the iterations if one of the following two conditions is fulfilled. The first condition (a) is:

$$\frac{\chi^2(\mathbf{x}_{i+1}) - \chi^2(\mathbf{x}_i)}{\chi^2(\mathbf{x}_{i+1})} < 0.001. \tag{4}$$

At the end of the retrievals we find $\chi^2 \approx \nu$ (see Table 1 of the discussion paper where we report χ^2/ν), therefore our condition is equivalent to:

$$\frac{\chi^2(\mathbf{x}_{i+1}) - \chi^2(\mathbf{x}_i)}{\nu} < 0.001 \tag{5}$$

i.e. it turns out that our first convergence criterion is more than conservative.

The second condition that we use (b) consists in stopping our retrievals after 10 iterations. Clearly this condition has no physical foundation and is motivated only by the need of limiting the computing time to a reasonable value. Most of the 1000 retrievals presented in the discussion paper terminate before 10 iterations, due to fulfilling of condition (a). In order to make sure that good physical convergence is also achieved by the retrievals terminating due to condition (b), for these retrievals we checked the value of the ratio $\left(\chi^2(\mathbf{x}_{10}) - \chi^2(\mathbf{x}_9)\right)/\chi^2(\mathbf{x}_{10})$. The maximum value for this ratio turns-out to be 0.029, i.e. the condition (3) suggested in Rodgers (2000) is still satisfied with a large margin.

ACPD
9, C10551–C10558, 2010

Let us now consider the concern of Dr. von Clarmann that, still far from convergence, a large LM damping λ_i could produce a very small χ^2 variation and erroneously trigger the convergence criterion (a). The largest value of λ_i that we obtain in the 1000 retrievals presented in the discussion paper is 6.4. This value occurs at the tenth iteration of some retrievals. We tried to carry-out one of these retrievals using $\lambda_i = 6.4$, constant through the iterations. We found that such a large λ_i is not able to trigger the convergence criterion (a) when the retrieval is still far from convergence. For this reason we conclude that the above mentioned concern does not apply to our retrievals.

Of course, since the LM term reduces the amplitude of the iteration step, the value of λ_i near the convergence has an impact on the final solution achieved. Therefore, we agree with Dr. von Clarmann that the LM damping has an effect on the result of the minimization. We argue, however, that this effect is not deleterious (false convergence far from the minimum) as he states, but, if properly used, is a tool to find a minimum of the cost function in ill-conditioned or ill-posed problems, where the GN method has no convergence. As mentioned in Doicu (2004) “for the conventional implementation (of the LM method) it is not clear what kind of stopping rule would be appropriate for ill-posed problems”. Therefore, despite of the conservativeness of our convergence criteria, these do not guarantee that the minimum of the cost function has been achieved with good accuracy. However, since the retrievals presented in the discussion paper operate on synthetic measurements, we have the possibility to check a-posteriori the accuracy of the obtained minimum. For this purpose we made the following two tests.

- **Statistics of χ^2/ν.** We calculated the average of the final values of χ^2/ν obtained in the 1000 retrievals presented in the discussion paper. We got 1.02, to be compared with the expected value of 1 (please remind that we are using simulated observations, therefore the forward model errors do not affect the χ^2). This difference is compatible with a convergence error at least a factor of 7 smaller than the retrieval error due to measurement noise.

- **Testing the differences between retrieved and true profiles**. We calculated the average $\overline{\alpha}$ over the 1000 retrievals of the quantity α defined as:

$$\alpha = \frac{(\mathbf{x}_\mathrm{r} - \mathbf{x}_\mathrm{true})^T \left(\mathbf{S}_\mathrm{r}^{(3)}\right)^{-1} (\mathbf{x}_\mathrm{r} - \mathbf{x}_\mathrm{true})}{n} \tag{6}$$

  where $\mathbf{x}_\mathrm{true}$ is the atmospheric profile assumed for the generation of simulated observations. For the other used symbols we refer to their definition contained in the discussion paper. If $\mathbf{S}_\mathrm{r}^{(3)}$ is the correct estimate of the retrieval error due to measurement noise and this is the only error component affecting the retrieved profiles, i.e if the convergence error is negligible, we expect $\alpha = 1$. Instead we get $\overline{\alpha} = 0.96$. This value indicates a negligible convergence error and is fully compatible with a 2% overestimation of the error in $\mathbf{S}_\mathrm{r}^{(3)}$. This marginal overestimation is consistent with the plot of Fig. 1 of the discussion paper.

We conclude that our retrievals reach a very good physical convergence and that the convergence error is negligible with respect to the retrieval error due to measurement noise.

As a final remark we point-out that, since we stop our retrievals as soon as condition (4) is satisfied, at the final iteration $\mathbf{x}_{i+1}$ and $\mathbf{x}_i$ differ by a fraction of their retrieval error but they are not equal within the machine precision. This means that, while our retrievals reach a very good “physical convergence”, the “exact numerical convergence” is not achieved. For this reason the foundational hypothesis of the proof contained in the first part of the SC, $\mathbf{x}_{i+1} = \mathbf{x}_i$, is not fulfilled by our retrievals. This is an additional reason why the conclusions of that proof do not apply to the results presented in the discussion paper.

The proof presented in the SC of Dr. von Clarmann applies only to well-conditioned retrievals that reached the “exact numerical convergence”, i.e. a quite “special” case

ACPD

9, C10551–C10558, 2010

in atmospheric retrievals. Conversely, the equations presented in the discussion paper provide accurate results in all cases: well- and ill-conditioned retrievals that reached either “physical” or the “exact numerical” convergence.

We plan to include in the revised paper part of the above results / discussion in order to avoid possible misunderstandings regarding the convergence of our retrievals.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 25663, 2009.

ACPD

9, C10551–C10558, 2010

Interactive Comment