Generalized least squares for spectral and dual energy CT: a simulation study

Cyril Mory, Bernhard Brendel, Klaus Erhard, Simon Rit

***Abstract*—In the presence of noise, decomposing spectral CT projections into materials generates anti-correlated noise. Estimating the covariance of this noise and taking it into account in the reconstruction process, by minimizing a GLS data-attachment term, is expected to lower the impact of the noise on the reconstruction. GLS has already been used in dual energy or spectral computed tomography, but always coupled with a regularization term, which raises the question of the relative impact of regularization and GLS. To our knowledge, a fair comparison between plain OLS and plain GLS is still missing. We provide one in this paper, with OLS and GLS reconstruction results from simulated projections, and discuss the relevance of using GLS. Pixels of the projection data are assumed to be independant, neglecting spatially correlated noise, and focusing on the inter-material noise correlation only. With these hypotheses, GLS brings little reduction of the noise level, while significantly increasing algorithmic complexity, slowing down convergence and requiring increased numerical precision with respect to OLS. Furthermore, in real situations, the covariance matrix has to be estimated, which adds another level of complexity and a potential source of inaccuracies.**

I. Introduction

In the presence of noise, decomposing spectral CT photon counts projections into material projections generates anti-correlated noise, since an over-estimation of the length traversed through one material must be compensated by an under-estimation of the length traversed through another to match the total attenuation. Estimating the covariance of this noise and taking it into account in the reconstruction process, by minimizing a Generalized Least Squares data-attachment term, theoretically guarantees that the estimate (the reconstructed volume) has the smallest possible variance, i.e. the least noise. GLS data-attachment terms taking into account the inter-materials correlation have already been used in dual energy [1] or spectral computed tomography [2] from material-decomposed projections, but since these studies use regularization, the impact of GLS alone cannot be evaluated from them. Throughout the paper, in order to avoid confusions, we use the term 'pixel' to describe an element of the projection data, and the term 'voxel' to describe an element of the reconstructed volume, even though the numerical experiments are actually 2-dimensional.

A. Principles of Ordinary and Generalized Least Squares

Ordinary least squares (OLS) and Generalized Least Squares (GLS) are two ways to design the cost function in an inverse problem. They do not dictate which algorithm should be used to perform minimization. When trying to retrieve a vector of parameters f from measurements p through a system of matrix R, OLS consists in minimizing $\|Rf - p\|_2^2$. If the errors on the measurements p are uncorrelated and of identical variance, minimizing the OLS cost function yields the Best Linear Unbiased Estimator (BLUE). If the errors on p are of different variances and/or correlated, the BLUE is obtained by minimizing the GLS cost function, i.e. $(Rf-p)^T C^{-1}(Rf-p)$, where C is the covariance matrix of p. Choosing $C = I$ yields the OLS cost function. OLS is therefore a specific case of GLS. In our case, the matrix R is the forward projection matrix, and f and p are column vectors representing 4D datasets: f is the set of 3D volumes to reconstruct, one per material, and p is the set of decomposed 2D material projections, one per X-ray source position and material. Throughout this document, we use the following dimension orders, from fastest to slowest:

- for f: material, then x, y, z axes of the volume
- for p: material, then u,v axes of the detector, then source position

B. Application to spectral CT

Since the decomposition method we used [3] processes each pixel of the projection data separately, it cannot provide spatial covariance information. The covariance matrix C is therefore restricted to inter-material covariance. With the dimension order specified in section I-A, it is block-diagonal, each block representing the $m \times m$ inter-material covariance matrix for a given pixel, where m is the number of materials. The forward projection matrix R is designed accordingly, and since all materials are projected the same way, it is made of blocks of m identical rows.

II. Material and methods

A. Minimization algorithm

Both OLS and GLS cost functions can be minimized by the linear conjugate gradient (CG) algorithm. CG solves problems of the type $Af = b$, where A is a symmetric positive definite matrix, f is the unknown vector and b is a known vector. Let us derive the expressions of A and b in the OLS and GLS cases:

- OLS: the minimum of the cost function is reached when its gradient is null, i.e. when

$$R^T(Rf - p) = 0$$
$$R^T Rf = R^T p$$

  Identifying the terms, $A = R^T R$ and $b = R^T p$. Note that if $R^T R$ can be inverted in a reasonable time, the Moore-Penrose pseudo inverse reconstruction $\widetilde{f} = (R^T R)^{-1} R^T b$ yields the same solution as a converged conjugate gradient initialized from zero.

C. Mory & S. Rit are with the Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1 ; Centre Léon Bérard, France.

B. Brendel & Klaus Erhard are with Philips GmbH Innovative Technologies, Research Laboratories, Hamburg, Germany

- GLS: let $U^TU = C^{-1}$ be the Cholesky decomposition of the inverse covariance matrix, with U an upper-triangular matrix. Then the GLS cost function can be rewritten as $\|U(Rf - p)\|_2^2$, and its gradient is null when

$$\begin{aligned} R^TU^TU(Rf - p) &= 0 \\ R^TU^TURf &= R^TU^TUp \\ R^TC^{-1}Rf &= R^TC^{-1}p \end{aligned}$$

  Identifying the terms, $A = R^TC^{-1}R$ and $b = R^TC^{-1}p$. The Moore-Penrose pseudo inverse reconstruction is $\widetilde{f} = (R^TC^{-1}R)^{-1}R^TC^{-1}b$.

The conjugate gradient algorithm is guaranteed to have converged after n iterations, where n is the number of elements in f. In practice, however, iterations are usually stopped far before that.

B. Simulation study

Simulations were performed with RTK [4]. We used a 2-dimensional, 3-materials (iodine, gadolinium and water) phantom, in which each material volume is made of 350×350 voxels. The phantom is composed of a small cylinder of iodine at concentration $1\,\text{mg.ml}^{-1}$, a small gadolinium cylinder at concentration $1\,\text{mg.ml}^{-1}$, both inside a large water cylinder at concentration $1\,\text{g.ml}^{-1}$, as shown in Fig. 1. 2400 projections of 643×1 pixels were computed analytically, using the geometry of the spectral CT scanner prototype (Philips Healthcare, Israël) installed at the CERMEP, Lyon, France. They were then converted to photon counts with 5 energy bins using the scanner's spectrum, detector response, and X-ray attenuation functions of each material. Poisson noise was added to the photon counts, then the noisy photon counts were decomposed into material projections using the RTK implementation of the method described in [3]. Since with noisy input photon count data, this method generates aberrant results for some pixels, we applied a Hampel filter [5] to remove the outliers. For each pixel, in each material, a $3 \times 3 \times 3$ neighborhood centered on the pixel is considered, and the median and standard deviation in this neighborhood are computed. If the pixel's value is off the median by more than 2σ, it is replaced by the median, otherwise it is left unchanged. In order to obtain a reliable estimate of the covariance of the noise on the decomposed projections, we repeated the last three steps (Poisson noise, decomposition and denoising) a hundred times, each time with a different realization of Poisson noise, and for each pixel, we computed the 3×3 covariance matrix on these hundred (3-materials) values. For the reconstructions, only one set of material projections was required, so we kept only the one from the first simulation.

Fig. 1: The three channels of the phantom used for simulations on RTK, shown side-by-side. From left to right: iodine, gadolinium, water

C. Evaluation of inter-material correlation coefficients

On the RTK simulated data, during the computation of the covariance matrix, we also computed the standard deviation of the noise in each material and each pixel, on the hundred realizations. Combining the covariance matrix and the standard deviation values, we were able to calculate correlations coefficients. Obviously, like the covariances, these vary from one pixel to the other, but the mean and standard deviation over all pixels is nevertheless interesting. The correlation coefficients we measured were the following:

- Iodine-gadolinium correlation's mean = -0.3963, std = 0.1185,
- Iodine-water correlation's mean = -0.1571, std = 0.1135,
- Gadolinium-water correlation's mean = -0.5489, std = 0.07953.

As expected, the noise is anti-correlated between materials.

D. Convergence issues and workarounds

While the mathematics of GLS are appealingly simple, we encountered a number of practical difficulties during our investigation of this topic. These difficulties are listed in the present section.

1) First CG iterates: It is well known that the first CG iterates of an OLS reconstruction are (roughly) low-pass approximations of the result CG yields at convergence, and high frequencies build up over the course of the iterations. In a GLS reconstruction, the first CG iterates are also low frequency, but they display a lot of cross-talk between materials (see Fig. 3). This cross-talk disappears over the course of the iterations. Stopping the iterations early is therefore much more problematic in GLS than in OLS. This problem can be alleviated by initializing the GLS conjugate gradient reconstruction with an early iterate of the corresponding OLS conjugate gradient reconstruction: with this approach, early stopping the GLS part no longer yields a result with a lot of cross-talk. However, it does not seem to reduce the number of iterations required to attain convergence. Therefore all results presented in this paper were obtained from a zero initialization.

2) Relative convergence speed: On the experiments we have carried out, GLS required many more iterations than OLS to reach convergence. Having to perform a lot of iterations for the GLS case brought in additional problems: after a certain point, each additional iteration of CG increased the noise in the reconstructions. This was due to the forward and a back projector not being the adjoint of one another with enough precision.

3) Adjoint operators: Having unmatched forward and back projectors is usually not a problem if the number of iterations remains low (typically below 100), but we had to do more iterations than that. We adopted the method described in [6], i.e. we computed the ratio between $< Rf, p >$ and $< f, R^Tp >$, where f and p are a random volume and a random set of projections, respectively, and $< ., . >$ denotes the dot product. The precision with which this ratio matches 1 is a measure of the 'adjointness' of R and R^T. In our case, using RTK's matched implementations of the Joseph forward and back projector [7], we obtained $1 - ratio \approx 10^{-5}$, which was insufficient. Switching from single precision (32-bits floats) to double precision

(64-bits) yielded $1 - ratio \approx 10^{-13}$, and got rid of the noise divergence problem. With our implementations, having to use matched forward and back projectors implies that at least one is sub-optimally implemented (either the ray-based back projector, like in our case, or a voxel-based forward projector, which is not available in RTK), which slows down calculations. In addition, having to use double precision both increases the reconstruction time again and doubles the memory requirements.

4) Scaling as preconditioning: Preconditioning consists in inserting a matrix D, and solving by CG the following problems instead of the ones described in section II-A:

- OLS:
$$DR^T RD^T \hat{f} = DR^T p$$
Identifying the terms, this means one must run the conjugate gradient algorithm with $A = DR^T RD^T$ and $b = DR^T p$, and obtain $\hat{f}$. Then $f = D^T \hat{f}$.
- GLS:
$$DR^T C^{-1} RD^T f = DR^T C^{-1} p$$
Identifying the terms, this means one must run the conjugate gradient algorithm with $A = DR^T C^{-1} RD^T$ and $b = DR^T C^{-1} p$, and obtain $\hat{f}$. Then, again, $f = D^T \hat{f}$.

Any matrix D can be used, but the goal of preconditioning is to obtain a new matrix A with a condition number closer to 1 than the original matrix A, because this condition number has a large impact on convergence speed (the closer it is to 1, the faster CG converges). In case of a diagonal preconditioner, $D = D^T$, and multiplying by D consists in a voxel-wise multiplication. We used

$$D = I_N \otimes \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

i.e. each material was simply scaled by a fixed factor λ_i. In our specific case, scaling can be performed either before or after forward projection, therefore $RD = D'R$ with D' a matrix similar to D but scaling projections instead of volumes. Defining $C' = D'C^{-1}D'$, the matrix A becomes $A = R^T C' R$ and the vector b becomes $b = R^T C' D'^{-1} p$. Preconditioning can therefore be obtained without increasing the amount of calculations at each iteration, simply by dividing the material projections by the $\lambda_i, i \in \{1..m\}$, performing a GLS reconstruction with a scaled covariance matrix C', and then multiplying the obtained material volumes by λ_i. We tried several choices for the λ_i: setting them all to 1, i.e. no preconditioning; setting them to the mean of the ground truth projection of each material; and $\lambda_i = \underset{E}{\text{mean}} \left(\frac{\mu_{iodine}(E)}{\mu_i(E)} \right)$, where $\mu_i(E)$ is the mass-attenuation of material i at energy E. The third method, although it does not depend of the object, and is therefore probably suboptimal in some cases, lead to the fastest convergence in our experiments.

5) Fair comparison: Because OLS and GLS converge at different speeds, choosing one iterate of each to perform a fair comparison between them is a delicate problem. The only possibility seems to be to wait until both have converged. After 1000 iterations, the GLS cost function reduction between successive iterates had dropped to 10^{-6}, which was deemed sufficient, while OLS had reached convergence after 150 iterations. Fig. 3 shows the convergence curves in logarithmic scale.

	OLS	GLS	Gain
MSE iodine	8.79e-07	7.85e-07	-10.7%
MSE gadolinium	4.62e-07	4.19e-07	-9.2%
MSE water	0.0352	0.0327	-7.0%
SNR iodine	0.478	0.486	+1.6%
SNR gadolinium	0.606	0.631	+4.1%
SNR water	2.497	2.572	+3.0%

TABLE I: Results on RTK simulations: compared MSEs and SNRs between OLS and GLS

III. Results

A. Photon counting experiments

Fig. 2 shows the results obtained with RTK, with one realization of the Poisson noise on photon counts. The GLS results are slightly less noisy than the OLS ones, but the improvement is only noticeable in the numerical analysis, and not visually. The MSE with respect to the ground truth and the SNR, computed as $\text{mean}(C)/\text{std}(C)$, where C is the set of voxels inside the cylinder of the material considered, are displayed in Table I.

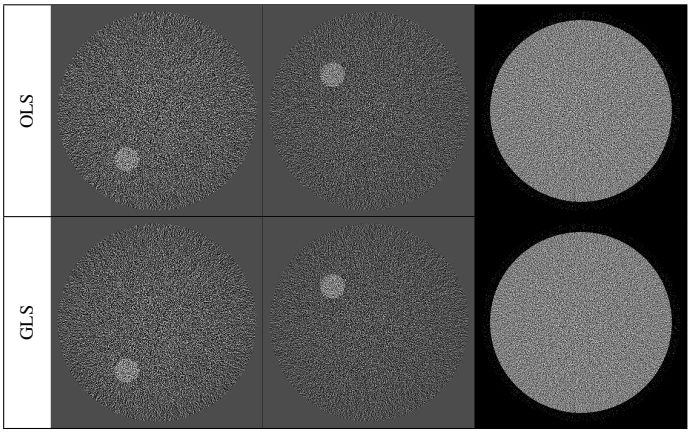

Fig. 2: RTK reconstructions with converged OLS and GLS. The GLS reconstructions are slightly less noisy, as shown in Table I, but it is not obvious on any of the reconstructions

B. Dual energy experiments

The small impact of GLS in the spectral CT experiments raises a question: could GLS be more effective in reducing the noise in a dual energy case, given that the anti-correlation between materials is much stronger in dual energy ? We ran the same RTK simulations as for spectral CT, only with a 2-materials phantom (water and iodine). The products of the detector responses by the incident spectrum were obtained by digitizing the green and pink curves in Fig. 1 of [8]. With this setup, we measured a mean correlation coefficient of -0.899. A Hampel filtering was applied on the material-decomposed projections, just like in spectral CT, and all subsequent results (covariance matrix, OLS and GLS reconstructions) were computed from the denoised projections. Despite the preconditioning, it took about 50 iterations for the OLS reconstruction to converge, and about 3000 iterations for the GLS reconstruction. Since, just like in the spectral case, the reconstructions OLS and GLS results are visually identical, they are not shown here, but image quality metrics are provided in Table II.

IV. Discussion

In an attempt to isolate how GLS improves over OLS from other contributions to noise reduction, we have chosen to perform unregularized reconstructions only, and have noticed

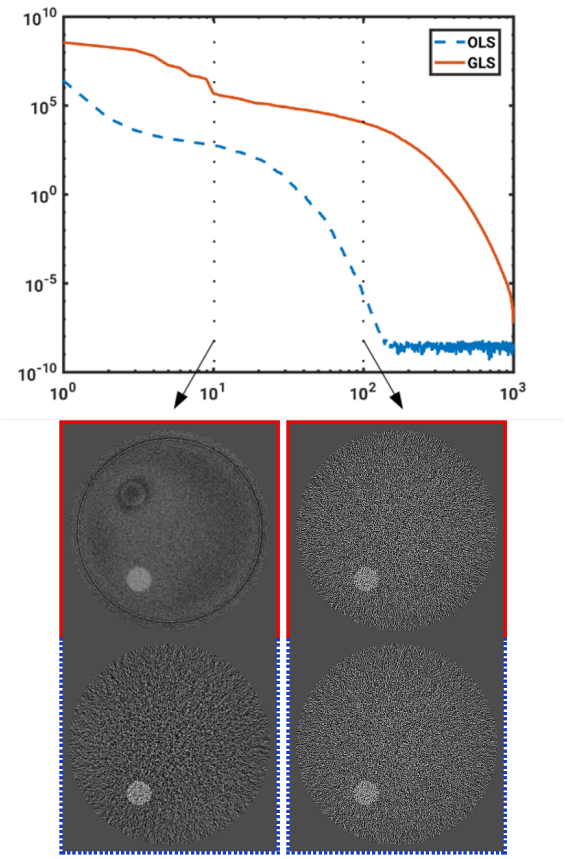

Fig. 3: Cost function after each iteration minus cost function after iteration 1000, in logarithmic scale. Iodine volumes at iterates 10 and 100 are shown below the graph for GLS and OLS.

	OLS	GLS	Gain
MSE iodine	9.18e-7	8.12e-7	-11.6%
MSE water	0.407	0.405	-0.57%
SNR iodine	0.442	0.470	+6.27%
SNR water	3.067	3.197	+4.25%

TABLE II: Results on RTK dual energy simulations: compared MSEs and SNRs between OLS and GLS

that GLS brought a very limited benefit. Whether this conclusion also holds for regularized reconstructions, and with which regularization method and parameters, remains an open question. Taking into account spatial correlations by accurately modeling advanced effects like scattering or charge sharing should increase the benefit of using GLS. However, it would also make implementation more complex, since C^{-1} would no longer be block-diagonal, therefore no longer "pixel-separable". In this work, the covariance matrix is estimated from a large number of repetitions of the same simulation (here 100), where only the Poisson noise realization changes. This empirical estimation is undoubtedly a source of inaccuracies, which could undermine the efficiency of GLS. To evaluate the importance of this error in our simulations, we ran a side experiment, adding noise with exactly the right covariance matrix (the one estimated by the simulations) to the noiseless projections, and reconstructing from those. This had no noticeable effect on the image quality of the GLS results, leading us to think that the limited impact of GLS is not caused by inaccuracies in the covariance matrix estimation. In addition, in a real situation, the covariance matrix estimation would most likely be less accurate than in the presented simulations.

V. Conclusion

Even with a quite precise knowledge of the covariance matrix, GLS only brings a moderate improvement of SNR in the RTK simulated case we have studied. The inevitable inaccuracies in estimating the covariance matrix in real situations are likely to reduce this improvement even more. Overall, the gain in SNR GLS can provide over OLS does not seem to be worth the efforts it requires in implementation, the risk that an incorrect covariance matrix might degrade the reconstruction, and the drop in performance and increase in memory footprint implied by matched projectors with double precision. Note that these conclusions hold for unregularized OLS and GLS: adding a regularizer may lead to different results.

Acknowledgements

This project was done in the framework of the EU's H2020 research and innovation program under the grant agreement No. 668142. This work was partially supported by grant ANR-17-CE19-0006-02 (ROIdoré project) from the French National Research Agency (ANR). This work was performed within the framework of the SIRIC LYric Grant INCa-DGOS-4664 and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

References

[1] Y. Liu, Z. Yu, and Y. Zou, "Impact of covariance modeling in dual-energy spectral CT image reconstruction," p. 94123Q, Mar. 2015.

[2] Q. Xu, A. Sawatzky, M. A. Anastasio, and C. O. Schirra, "Sparsity-regularized image reconstruction of decomposed K-edge data in spectral CT," *Physics in Medicine and Biology*, vol. 59, pp. N65–N79, May 2014.

[3] J. P. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G Zeitler, A. Livne, and R. Proksa, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine and Biology*, vol. 53, no. 15, p. 4031, 2008.

[4] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp, "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)," in *Proceedings of the International Conference on the Use of Computers in Radiation Therapy (ICCR)*, 2013.

[5] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, "Generalized Hampel Filters," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, p. 87, Dec. 2016.

[6] F. Arcadu, M. Stampanoni, and F. Marone, "On the crucial impact of the coupling projector-backprojector in iterative tomographic reconstruction," *arXiv:1612.05515 [cs]*, Dec. 2016. arXiv: 1612.05515.

[7] P. M. Joseph, "An Improved Algorithm for Reprojecting Rays through Pixel Images," *Medical Imaging, IEEE Transactions on*, no. 3, pp. 192 – 196, 1982.

[8] K. Brown, S. Zabic, and G. Shechter, "Impact of Spectral Separation in Dual-Energy CT with Anti-Correlated Statistical Reconstruction," pp. 491–494, 2015.