

Bayesian inference with error variable splitting and sparsity enforcing priors for linear inverse problems

Ali Mohammad-Djafari, Mircea Dumitru, Camille Chapdelaine,
Nicolas Gac

Laboratoire des Signaux et Systèmes (L2S)
UMR8506 CNRS-CentraleSupélec-UNIV PARIS SUD
SUPELEC, 91192 Gif-sur-Yvette, France

Email: djafari@lss.supelec.fr
<http://djafari.free.fr>

EUSIPCO 2018, Rome, Italy

Linear Inverse Problems: Regularization methods

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon},$$

- ▶ Regularization (L2, L1, TV, ...)

$$J(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda R(\mathbf{f})$$

$$R(\mathbf{f}) = \left\{ \|\mathbf{f}\|_2^2, \|\mathbf{f}\|_\beta^\beta, \|\mathbf{D}\mathbf{f}\|_2^2, \|\mathbf{D}\mathbf{f}\|_\beta^\beta, \sum_j \phi([\mathbf{D}\mathbf{f}]_j) \right\}, \quad 1 \leq \beta \leq 2$$

- ▶ Optimization algorithms

- ▶ $R(\mathbf{f})$ quadratic: Gradient based, CG
- ▶ $R(\mathbf{f})$ nonquadratic, but convex and derivable: Gradient based, Conjugate Gradient (CG)
- ▶ $R(\mathbf{f})$ convex but non-derivable: sub-gradient
- ▶ $R(\mathbf{f})$ convex: all the convex optimization algorithms

Linear Inverse Problems: L2 & L1 Regularization methods

- ▶ L2 or quadratic: $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon,$

$$J(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{D}\mathbf{f}\|_2^2$$

Analytic solution: $\hat{\mathbf{f}} = (\mathbf{H}'\mathbf{H} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{H}'\mathbf{g}$

Gradient based: $\nabla J(\mathbf{f}) = -\mathbf{H}'(\mathbf{g} - \mathbf{H}\mathbf{f}) + 2\lambda\mathbf{D}'\mathbf{D}\mathbf{f}$

- ▶ L1 : convex but not derivable at zero:

$$J(\mathbf{f}) = \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{D}\mathbf{f}\|_1$$

Fenchel conjugate, Dual problem, subgradient, proximal operator, ...

- ▶ Variable splitting and Augmented Lagrangian

$$(\mathbf{f}, \hat{\mathbf{z}}) = \arg \min_{(\mathbf{f}, \mathbf{z})} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda \|\mathbf{z}\|_1 + q \|\mathbf{z}\|_2^2 \right\} \quad \text{s.t.} \quad \mathbf{z} = \mathbf{D}\mathbf{f}$$

- ▶ A great number of optimization algorithms have been proposed: ADMM, ISTA, FISTA, ...

Linear Inverse Problems: Bayesian approach

- ▶ Bayese rule $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon,$

$$p(\mathbf{f}|\mathbf{g}) = \frac{p(\mathbf{g}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})} \text{ with } p(\mathbf{g}) = \int p(\mathbf{g}|\mathbf{f})p(\mathbf{f}) d\mathbf{f}$$

- ▶ Gaussian noise:

$$p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_\epsilon \mathbf{I}) \propto \exp \left[\frac{-1}{2v_\epsilon} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 \right]$$

- ▶ Prior models:

$$p(\mathbf{f}) \propto \exp \left[\frac{-1}{2v_f} \|\mathbf{f}\|_2^2 \right], \quad \exp \left[\frac{-1}{2v_f} \|\mathbf{D}\mathbf{f}\|_2^2 \right], \quad \exp \left[\frac{-1}{2v_f} \|\mathbf{f}\|_1 \right], \dots$$

- ▶ MAP estimate: $J(\mathbf{f}) = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \lambda R(\mathbf{f})$:

$$p(\mathbf{f}|\mathbf{g}) \propto \exp \left[\frac{-1}{2v_\epsilon} J(\mathbf{f}) \right] \rightarrow \hat{\mathbf{f}}_{MAP} = \arg \max_{\mathbf{f}} \{p(\mathbf{f}|\mathbf{g})\} = \arg \min_{\mathbf{f}} \{J(\mathbf{f})\}$$

State of the Art Bayesian methods for the simple case

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$$

Supervised Gaussian case:

$$\begin{cases} p(\mathbf{g}|\mathbf{f}) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_{\epsilon}\mathbf{I}) \\ p(\mathbf{f}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, v_f\mathbf{I}) \end{cases} \rightarrow \begin{cases} p(\mathbf{f}|\mathbf{g}) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}) \\ \hat{\mathbf{f}} = [\mathbf{H}'\mathbf{H} + \lambda\mathbf{I}]^{-1}\mathbf{H}'\mathbf{g} \\ \hat{\boldsymbol{\Sigma}} = v_{\epsilon}[\mathbf{H}'\mathbf{H} + \lambda\mathbf{I}]^{-1}, \quad \lambda = \frac{v_{\epsilon}}{v_f} \end{cases}$$

Unsupervised Gaussian

$$\begin{cases} p(\mathbf{g}|\mathbf{f}, v_{\epsilon}) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, v_{\epsilon}\mathbf{I}) \\ p(\mathbf{f}|v_f) = \mathcal{N}(\mathbf{f}|\mathbf{0}, v_f\mathbf{I}) \\ p(v_{\epsilon}) = \mathcal{IG}(v_{\epsilon}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\ p(v_f) = \mathcal{IG}(v_f|\alpha_{f_0}, \beta_{f_0}) \end{cases} \rightarrow \begin{cases} p(\mathbf{f}|\mathbf{g}, v_{\epsilon}, v_f) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, \hat{\boldsymbol{\Sigma}}) \\ \hat{\mathbf{f}} = [\mathbf{H}'\mathbf{H} + \hat{\lambda}\mathbf{I}]^{-1}\mathbf{H}'\mathbf{g} \\ \hat{\boldsymbol{\Sigma}} = \hat{v}_{\epsilon}[\mathbf{H}'\mathbf{H} + \hat{\lambda}\mathbf{I}]^{-1}, \quad \hat{\lambda} = \frac{\hat{v}_{\epsilon}}{v_f} \\ p(v_{\epsilon}|\mathbf{g}, \mathbf{f}) = \mathcal{IG}(v_{\epsilon}|\tilde{\alpha}_{\epsilon}, \tilde{\beta}_{\epsilon}) \\ p(v_f|\mathbf{g}, \mathbf{f}) = \mathcal{IG}(v_f|\tilde{\alpha}_f, \tilde{\beta}_f) \\ \tilde{\alpha}_{\epsilon}, \tilde{\beta}_{\epsilon}, \tilde{\alpha}_f, \tilde{\beta}_f \end{cases}$$

$$p(\mathbf{f}, v_\epsilon, v_\zeta | \mathbf{g}) \propto \exp[-J(\mathbf{f}, v_\epsilon, v_\zeta)]$$

- ▶ JMAP: Alternate optimization with respect to \mathbf{f} , v_ϵ , v_f :

$$J(\mathbf{f}, v_\epsilon, v_f) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \frac{1}{2v_f} \|\mathbf{f}\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon} + (\alpha_{f_0} + 1) \ln v_f + \frac{\beta_{f_0}}{v_f}$$

- ▶ Gibbs sampling MCMC:

$$\mathbf{f} \sim p(\mathbf{f}, v_\epsilon, v_\zeta | \mathbf{g}) \rightarrow v_\epsilon \sim p(v_\epsilon | \mathbf{g}, \mathbf{f}) \rightarrow v_f \sim p(v_f | \mathbf{g}, \mathbf{f})$$

- ▶ Variational Bayesian Approximation: Approximate $p(\mathbf{f}, v_\epsilon, v_\zeta | \mathbf{g})$ by a separable one $q(\mathbf{f}, v_\epsilon, v_\zeta) = q_1(\mathbf{f})q_2(v_\epsilon)q_3(v_\zeta)$ minimizing $\text{KL}(q|p)$.

Hierarchical models for more robustness

$$\begin{cases} \mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}, \\ \mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\zeta}, \end{cases} \quad \mathbf{z} \text{ sparse Student} \rightarrow \begin{cases} p(\mathbf{z}_j | \mathbf{v}_{zj}) = \mathcal{N}(\mathbf{z}_j | 0, \mathbf{v}_{zj}), \\ p(\mathbf{v}_{zj}) = \mathcal{IG}(\mathbf{v}_{zj} | \alpha_{z_0}, \beta_{z_0}) \end{cases}$$

$$\begin{cases} p(\mathbf{g} | \mathbf{f}) = \mathcal{N}(\mathbf{g} | \mathbf{H}\mathbf{f}, \mathbf{v}_\epsilon \mathbf{I}) \\ p(\mathbf{f} | \mathbf{z}) = \mathcal{N}(\mathbf{f} | \mathbf{D}\mathbf{z}, \mathbf{v}_\zeta \mathbf{I}) \\ p(\mathbf{z} | \mathbf{v}_z) = \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbf{V}_z) \rightarrow \\ p(\mathbf{v}_z) = \prod_j \mathcal{IG}(\mathbf{v}_{zj} | \alpha_{z_0}, \beta_{z_0}) \\ p(\mathbf{v}_\epsilon) = \mathcal{IG}(\mathbf{v}_\epsilon | \alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\ p(\mathbf{v}_\zeta) = \mathcal{IG}(\mathbf{v}_\zeta | \alpha_{\zeta_0}, \beta_{\zeta_0}) \end{cases} \left\{ \begin{array}{l} p(\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_\zeta | \mathbf{g}) \propto \exp[-J(\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_\zeta)] \\ J(\mathbf{f}, \mathbf{z}, \mathbf{v}_z, \mathbf{v}_\epsilon, \mathbf{v}_\zeta) = \frac{1}{2\mathbf{v}_\epsilon} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|_2^2 + \\ \frac{1}{2\mathbf{v}_\zeta} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2 + \|\mathbf{V}_z^{-\frac{1}{2}} \mathbf{z}\|_2^2 + \\ \sum_j (\alpha_{z_0} + 1) \ln \mathbf{v}_{zj} + \beta_{z_0} / \mathbf{v}_{zj} + \\ (\alpha_{\epsilon_0} + m/2) \ln \mathbf{v}_\epsilon + \beta_{\epsilon_0} / \mathbf{v}_\epsilon + \\ (\alpha_{\zeta_0} + n/2) \ln \mathbf{v}_\zeta + \beta_{\zeta_0} / \mathbf{v}_\zeta \end{array} \right.$$

Main advantages:

- ▶ Quadratic optimization with respect to \mathbf{f} and \mathbf{z}
- ▶ Direct updates of the hyperparameters \mathbf{v}_ϵ and \mathbf{v}_ζ
- ▶ Possibility to compute posterior mean and quantify uncertainties analytically via VBA

Variable splitting or How to account for all uncertainties

Standard case:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon \rightarrow \begin{cases} \mathbf{g} = \mathbf{g}_0 + \epsilon, \\ \mathbf{g}_0 = \mathbf{H}\mathbf{f} \end{cases}$$

Error splitting:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta} + \epsilon \rightarrow \begin{cases} \mathbf{g} = \mathbf{g}_0 + \epsilon, & \text{measurement error} \\ \mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta}, & \text{modeling error} \end{cases}$$

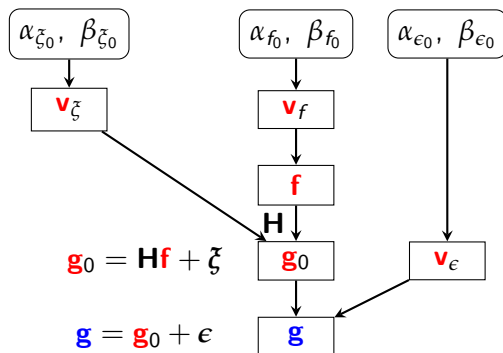
or even

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{u} + \boldsymbol{\zeta} + \epsilon \rightarrow \begin{cases} \mathbf{g} = \mathbf{g}_0 + \mathbf{u} + \epsilon, & \mathbf{u} \text{ background} \\ \mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta} \end{cases}$$

Proposed Error splitting Method 1

$$\begin{cases} \mathbf{g} = \mathbf{g}_0 + \boldsymbol{\epsilon}, & \text{measurement error: Gaussian with unknown variance} \\ \mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta}, & \text{modeling error: Student-t or Normal-Gamma} \end{cases}$$

- ▶ Direct Sparsity enforcing prior: $(\mathbf{f}, \mathbf{v}_f)$ Normal-IG, \mathbf{f} Student-t



Proposed Error splitting Method 1 (details)

- ▶ $\mathbf{g} = \mathbf{g}_0 + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}$ Gaussian:

$$p(\mathbf{g}|\mathbf{g}_0, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{g}_0, \mathbf{v}_\epsilon \mathbf{I}), \quad p(\mathbf{v}_\epsilon) = \mathcal{IG}(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}),$$

- ▶ $\mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta}$, $\boldsymbol{\zeta}$ Student:

$$\begin{cases} p(\mathbf{g}_0|\mathbf{f}, \mathbf{v}_\zeta) = \mathcal{N}(\mathbf{g}_0|\mathbf{H}\mathbf{f}, \mathbf{V}_\zeta), & \mathbf{V}_\zeta = \text{diag}\{\mathbf{v}_\zeta\}, \\ p(\mathbf{v}_\zeta) = \prod_{i=1}^M p(\mathbf{v}_{\zeta i}) = \prod_{i=1}^M \mathcal{IG}(\mathbf{v}_{\zeta i}|\alpha_{\zeta_0}, \beta_{\zeta_0}), \end{cases}$$

- ▶ \mathbf{f} Sparse:

$$\begin{cases} p(\mathbf{f}|\mathbf{v}_f) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{V}_f), & \mathbf{V}_f = \text{diag}\{\mathbf{v}_f\} \\ p(\mathbf{v}_f) = \prod_{j=1}^N p(\mathbf{v}_{f j}) = \prod_{j=1}^N \mathcal{IG}(\mathbf{v}_{f j}|\alpha_{f_0}, \beta_{f_0}) \end{cases}$$

which results in:

$$p(\mathbf{f}, \mathbf{g}_0, \mathbf{v}_\epsilon, \mathbf{v}_\zeta, \mathbf{v}_f|\mathbf{g}) \propto \exp[-J(\mathbf{f}, \mathbf{g}_0, \mathbf{v}_\epsilon, \mathbf{v}_\zeta, \mathbf{v}_f)] \quad \text{with}$$

$$\begin{aligned}
 J(\mathbf{f}, \mathbf{g}_0, v_\epsilon, \mathbf{v}_{\zeta}, \mathbf{v}_f) &= \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon} \\
 &+ \frac{1}{2} \|\mathbf{V}_{\zeta}^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \sum_{i=1}^M \left[(\alpha_{\zeta_0} + 1) \ln v_{\zeta_i} + \frac{\beta_{\zeta_0}}{v_{\zeta_i}} \right] \\
 &+ \frac{1}{2} \|\mathbf{V}_f^{-1/2} \mathbf{f}\|_2^2 + \sum_{j=1}^N \left[(\alpha_{f_0} + 1) \ln v_{f_j} + \frac{\beta_{f_0}}{v_{f_j}} \right]
 \end{aligned}$$

Alternate optimization:

▶ with respect to \mathbf{f} : $J(\mathbf{f}) = \frac{1}{2} \|\mathbf{V}_{\zeta}^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \frac{1}{2v_{\zeta}} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2$

▶ with respect to \mathbf{g}_0 :

$$J(\mathbf{g}_0) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + \frac{1}{2} \|\mathbf{V}_{\zeta}^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2$$

▶ with respect to v_ϵ : $J(v_\epsilon) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon}$

▶ with respect to v_{ζ_i} :

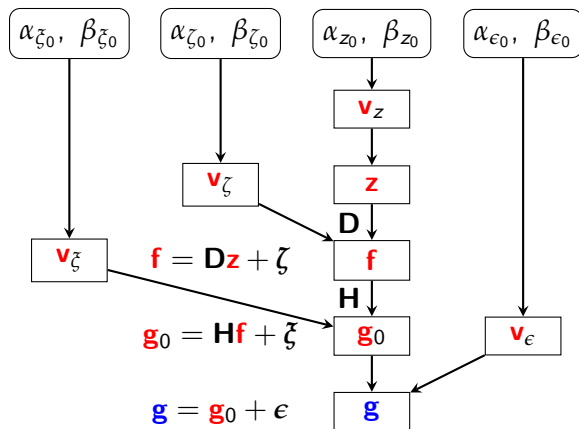
$$J(v_{\zeta_i}) = \frac{1}{2} \|\mathbf{V}_{\zeta}^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \sum_{i=1}^M \left[(\alpha_{\zeta_0} + 1) \ln v_{\zeta_i} + \frac{\beta_{\zeta_0}}{v_{\zeta_i}} \right]$$

▶ with respect to v_{f_j} :

$$J(v_{f_j}) = \frac{1}{2} \|\mathbf{V}_z^{-1/2} \mathbf{f}\|_2^2 + \sum_{j=1}^N \left[(\alpha_{f_0} + 1) \ln v_{f_j} + \frac{\beta_{f_0}}{v_{f_j}} \right]$$

Proposed Error splitting Method 2

$$\left\{ \begin{array}{l} \mathbf{g} = \mathbf{g}_0 + \boldsymbol{\epsilon}, \quad \text{measurement error: Gaussian with unknown variance} \\ \mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta}, \quad \text{modeling error: Student-t or Normal-Gamma} \\ \mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\zeta}, \quad \text{Sparsity enforcing in Transformed domain} \end{array} \right.$$



Proposed Error splitting Method 2 (details)

- ▶ $\mathbf{g} = \mathbf{g}_0 + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}$ Gaussian:

$$p(\mathbf{g}|\mathbf{g}_0, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{g}_0, \mathbf{v}_\epsilon \mathbf{I}), \quad p(\mathbf{v}_\epsilon) = \mathcal{IG}(\mathbf{v}_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}),$$

- ▶ $\mathbf{g}_0 = \mathbf{H}\mathbf{f} + \boldsymbol{\zeta}$, $\boldsymbol{\zeta}$ Student:

$$\begin{cases} p(\mathbf{g}_0|\mathbf{f}, \mathbf{v}_\zeta) = \mathcal{N}(\mathbf{g}_0|\mathbf{H}\mathbf{f}, \mathbf{V}_\zeta), & \mathbf{V}_\zeta = \text{diag}\{\mathbf{v}_\zeta\}, \\ p(\mathbf{v}_\zeta) = \prod_{i=1}^M p(v_{\zeta i}) = \prod_{i=1}^M \mathcal{IG}(v_{\zeta i}|\alpha_{\zeta_0}, \beta_{\zeta_0}), \end{cases}$$

- ▶ $\mathbf{f} = \mathbf{D}\mathbf{z} + \boldsymbol{\zeta}$, $\boldsymbol{\zeta}$ Gaussian:

$$p(\mathbf{f}|\mathbf{z}, \mathbf{v}_\zeta) = \mathcal{N}(\mathbf{f}|\mathbf{D}\mathbf{z}, \mathbf{v}_\zeta \mathbf{I}), \quad p(\mathbf{v}_\zeta) = \mathcal{IG}(\mathbf{v}_\zeta|\alpha_{\zeta_0}, \beta_{\zeta_0}),$$

- ▶ \mathbf{z} Sparse:

$$\begin{cases} p(\mathbf{z}|\mathbf{v}_z) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{V}_z), & \mathbf{V}_z = \text{diag}\{\mathbf{v}_z\} \\ p(\mathbf{v}_z) = \prod_{j=1}^N p(v_{zj}) = \prod_{j=1}^N \mathcal{IG}(v_{zj}|\alpha_{z_0}, \beta_{z_0}) \end{cases}$$

which results in:

$$p(\mathbf{f}, \mathbf{g}_0, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\zeta, \mathbf{v}_z|\mathbf{g}) \propto \exp[-J(\mathbf{f}, \mathbf{g}_0, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_\zeta, \mathbf{v}_z)] \quad \text{with}$$

$$\begin{aligned}
J(\mathbf{f}, \mathbf{g}_0, \mathbf{z}, v_\epsilon, v_\zeta, \mathbf{v}_z) &= \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon} \\
&+ \frac{1}{2} \|\mathbf{V}_\zeta^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \sum_{i=1}^M \left[(\alpha_{\zeta_0} + 1) \ln v_{\zeta_i} + \frac{\beta_{\zeta_0}}{v_{\zeta_i}} \right] \\
&+ \frac{1}{2v_\zeta} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2 + (\alpha_{\zeta_0} + 1) \ln v_\zeta + \frac{\beta_{\zeta_0}}{v_\zeta} \\
&+ \frac{1}{2} \|\mathbf{V}_z^{-1/2} \mathbf{z}\|_2^2 + \sum_{j=1}^N \left[(\alpha_{z_0} + 1) \ln v_{z_j} + \frac{\beta_{z_0}}{v_{z_j}} \right]
\end{aligned}$$

Alternate optimization:

▶ with respect to \mathbf{f} : $J(\mathbf{f}) = \frac{1}{2} \|\mathbf{V}_\zeta^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \frac{1}{2v_\zeta} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2$

▶ with respect to \mathbf{g}_0 :

$$J(\mathbf{g}_0) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + \frac{1}{2} \|\mathbf{V}_\zeta^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2$$

▶ with respect to \mathbf{z} : $J(\mathbf{z}) = \frac{1}{2v_\zeta} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2 + \frac{1}{2} \|\mathbf{V}_z^{-1/2} \mathbf{z}\|_2^2$

▶ with respect to v_ϵ : $J(v_\epsilon) = \frac{1}{2v_\epsilon} \|\mathbf{g} - \mathbf{g}_0\|_2^2 + (\alpha_{\epsilon_0} + 1) \ln v_\epsilon + \frac{\beta_{\epsilon_0}}{v_\epsilon}$

▶ v_{ζ_i} : $J(v_{\zeta_i}) = \frac{1}{2} \|\mathbf{V}_\zeta^{-1/2} (\mathbf{g}_0 - \mathbf{H}\mathbf{f})\|_2^2 + \sum_{i=1}^M \left[(\alpha_{\zeta_0} + 1) \ln v_{\zeta_i} + \frac{\beta_{\zeta_0}}{v_{\zeta_i}} \right]$

▶ with respect to v_ζ : $J(v_\zeta) = \frac{1}{2v_\zeta} \|\mathbf{f} - \mathbf{D}\mathbf{z}\|_2^2 + (\alpha_{\zeta_0} + 1) \ln v_\zeta + \frac{\beta_{\zeta_0}}{v_\zeta}$

▶ to v_{z_j} : $J(v_{z_j}) = \frac{1}{2} \|\mathbf{V}_z^{-1/2} \mathbf{z}\|_2^2 + \sum_{j=1}^N \left[(\alpha_{z_0} + 1) \ln v_{z_j} + \frac{\beta_{z_0}}{v_{z_j}} \right]$

Main advantages and limitations

Advantages:

- ▶ All the optimization are either quadratic or explicite
- ▶ Quadratic optimizations can be done efficiently
- ▶ For great dimensional problems, the needed operators \mathbf{H} , \mathbf{H}' , \mathbf{D} and \mathbf{D}' can be implemented on GPU
- ▶ For Computed Tomography, efficient GPU implementation of these operators have been done in our group for 2D and 3D CT.

Limitations:

- ▶ Huge amount of memory is needed for \mathbf{f} , \mathbf{g}_0 , \mathbf{v}_ζ and \mathbf{v}_z
- ▶ No easy way to study the global convergency of the algorithm.
- ▶ The number of hyper-hyperparameters $(\alpha_{\epsilon_0}, \beta_{\epsilon_0})$, $(\alpha_{\zeta_0}, \beta_{\zeta_0})$, $(\alpha_{z_0}, \beta_{z_0})$ to be fixed is important.

Applications:

- ▶ Biological time series analysis: Estimating periodic components in very short time biological time series
- ▶ Signal deconvolution in Mass Spectrometry
- ▶ Image restoration in RAMAN Mass Spectrometry or in Radio Astronomy
- ▶ 2D and 3D Industrial Computed Tomography for Non Destructive Testing Application
- ▶ Low dose and limited angle CT for biological or medical applications