E9 203 Compressed Sensing & Sparse SP

05 Apr. 2021

Last time:

Q. Why does SBL lead to sparse solns?

Ans 1. Impose an inverse gamma hyperprior on γ:

$$\Rightarrow p_\gamma(\gamma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{-\alpha-1} e^{-\beta/\gamma}, \quad \gamma \geq 0,\ \alpha, \beta \geq 0.$$

$$\Rightarrow \boxed{p_x(x) \propto \prod_{i=1}^{N} \left(b + \frac{x_i^2}{2}\right)^{-(a+\frac{1}{2})}} \xrightarrow{a,b\to 0} \prod_{i=1}^{N} \frac{1}{|x_i|} \text{ : encourages sparsity.}$$

Ans. 2: Instead of working with $p_x(x)$, define a family of approx. to $p_x(x)$, and pick the member of the family that maximizes Bayesian evidence.

Letting $z_i = x_i^2$, and
$$f(z_i) \triangleq \log p_x(x_i) = -\left(a+\tfrac{1}{2}\right)\log\left(b + \frac{z_i}{2}\right) + \text{const.}$$

$f(z_i)$: convex fn. of $z_i \Rightarrow$ using the dual form

$$\begin{cases} f(z) = \sup_\lambda \ \lambda z - f^*(\lambda) \\ f^*(\lambda) \triangleq \sup_z \ \lambda z - f(z) \quad \text{convex dual fn.} \end{cases}$$

We showed that
$$p_x(x) \propto \prod_{i=1}^{N} \max_{\gamma_i \geq 0} \frac{1}{\sqrt{2\pi\gamma_i}} \gamma_i^{-a} e^{-\frac{b}{\gamma_i}} e^{-\frac{x_i^2}{2\gamma_i}}$$

As $a, b \to 0$, this becomes
$$p_x(x) \propto \prod_{i=1}^{N} \max_{\gamma_i \geq 0} \underbrace{\frac{1}{\sqrt{2\pi\gamma_i}} e^{-\frac{x_i^2}{2\gamma_i}}}_{\text{Parametrized set of approx. to } p_x(x)\,:\ \mathcal{N}(0,\Gamma).}$$

Thus, the SBL procedure considers a variational approx to a sparsity-promoting prior $p_x(x) \propto \prod_{i=1}^{N}\left(b + \frac{x_i^2}{2}\right)^{-(a+\frac{1}{2})}$, and finds the "tightest" lower bound via type-II ML.

McKay Update (see also: Faul & Tipping, "Analysis of SBL".)

SBL cost fn.:
$$L(\gamma) = -\log p(y;\gamma) = \log\det(\sigma^2 I + A\Gamma A^T) + y^T(\sigma^2 I + A\Gamma A^T)^{-1} y$$

Convenient to work with inverse variances: $\alpha_i = \frac{1}{\gamma_i}$, $i \in [N]$.

Let $\mathcal{A} = \operatorname{diag}([\alpha_1 \ldots \alpha_N]^T)$

$$C = \sigma^2 I + A\mathcal{A}^{-1}A^T$$

Then $\mathcal{L}(\alpha) = \log\det C + y^T C^{-1} y$: to minimize wrt α.

Write
$$C = \sigma^2 I + \sum_{\ell=1}^{N} \alpha_\ell^{-1} a_\ell a_\ell^T$$
$$= \sigma^2 I + \sum_{\ell \neq i} \alpha_\ell^{-1} a_\ell a_\ell^T + \alpha_i^{-1} a_i a_i^T$$
$$= \underbrace{C_{-i}}_{\triangleq\, \sigma^2 I + \sum_{\ell\neq i} \alpha_\ell^{-1} a_\ell a_\ell^T} + \alpha_i^{-1} a_i a_i^T$$

Determinant and inverse identities:

$$\begin{cases} \det C = (\det C_{-i})(1 + \alpha_i^{-1} a_i^T C_{-i}^{-1} a_i) \\ C^{-1} = C_{-i}^{-1} - \dfrac{C_{-i}^{-1} a_i a_i^T C_{-i}^{-1}}{\alpha_i + a_i^T C_{-i}^{-1} a_i}. \quad \text{[Rank-1 update formula]} \end{cases}$$

$$\Rightarrow \mathcal{L}(\alpha) = \log\det C_{-i} + y^T C_{-i}^{-1} y - \log\alpha_i + \log(\alpha_i + a_i^T C_{-i}^{-1} a_i) - \frac{(a_i^T C_{-i}^{-1} y)^2}{\alpha_i + a_i^T C_{-i}^{-1} a_i}$$

$$= \underbrace{\mathcal{L}(\alpha_{-i})}_{\text{log marginal likelihood with } \alpha_i \text{ removed.}} + \ell(\alpha_i)$$

$\ell(\alpha_i)$ isolates the terms involving α_i.

Differentiating $\mathcal{L}(\alpha)$ w.r.t. α_i,

$$\frac{\partial \mathcal{L}(\alpha)}{\partial \alpha_i} = \frac{\partial \ell(\alpha_i)}{\partial \alpha_i} = \frac{-1}{\alpha_i} + \frac{1}{\alpha_i + a_i^T C_{-i}^{-1} a_i} + \frac{(a_i^T C_{-i}^{-1} y)^2}{(\alpha_i + a_i^T C_{-i}^{-1} a_i)^2}$$

$$= \frac{-\alpha_i^{-1} s_i^2 + (Q_i^2 - s_i)}{(\alpha_i + s_i)^2}$$

where $Q_i \triangleq a_i^T C_{-i}^{-1} y$, $\quad s_i \triangleq a_i^T C_{-i}^{-1} a_i$

Setting the derivative $= 0$,
$$\alpha_i = \frac{s_i^2}{Q_i^2 - s_i} \quad \text{if } Q_i^2 > s_i \qquad y = Ax + w$$

(and can s.t. $\alpha_i = \infty$ o/w, see paper)

(also, when $Q_i^2 - s_i > 0$, α_i is +ve as desired)

The update of α_i is inexpensive to compute: C_{-i}^{-1} can be computed fast using the matrix inversion lemma/rank-1 update formula (see above).

Reweighted ℓ_1 approach for SBL cost min.

$$\mathcal{L}(\gamma) = \log\det\Sigma_y + y^T\Sigma_y^{-1}y \text{ : min wrt } \gamma.$$

where $\Sigma_y = \sigma^2 I + A\Gamma A^T$

$\log\det\Sigma_y$ is concave in Γ. Use concave dual repn.:

$$\log\det\Sigma_y = \min_z \ z^T\gamma - g^*(z) \quad \text{— ①}$$

$$g^*(z) \triangleq \min_\gamma \ z^T\gamma - \log\det\Sigma_y \quad \text{— ②}$$

$$\Rightarrow L(\gamma, z) = z^T\gamma - g^*(z) + y^T\Sigma_y^{-1}y \geq L(\gamma)$$

Alternately min. $L(\gamma, z)$ wrt γ & z.

Given γ, update for z comes from ②:
$$z_{opt} = \nabla_\gamma \log\det\Sigma_y$$

Given z, update for γ:
$$\gamma \leftarrow \arg\min_\gamma \ L_z(\gamma) = z^T\gamma + y^T\Sigma_y^{-1}y$$

Now, $y^T\Sigma_y^{-1}y = \min_x \frac{1}{\sigma^2}\|y - Ax\|_2^2 + \sum_{i=1}^{N}\frac{x_i^2}{\gamma_i}$ [HW]

$\Rightarrow$ Auxiliary upper bound:
$$L_z(\gamma, x) = \sum_{i=1}^{N} z_i\gamma_i + \frac{x_i^2}{\gamma_i} + \frac{1}{\sigma^2}\|y - Ax\|_2^2 \geq L_z(\gamma)$$

Jointly convex in γ and x. Minimizing wrt γ, we get
$$\gamma_i = z_i^{-1/2}|x_i|.$$

Substituting,
$$L_z(\gamma_{opt}, x) = \sum_{i=1}^{N}\sqrt{z_i}\,|x_i| + \sqrt{z_i}\,|x_i| + \frac{1}{\sigma^2}\|y - Ax\|_2^2$$

$\Rightarrow \gamma_{opt}$ can be found by solving
$$x_* = \arg\min_x \ \|y - Ax\|_2^2 + 2\sigma^2\sum_{i=1}^{N}\sqrt{z_i}\,|x_i| \quad \text{— (*)}$$
[Weighted ℓ_1 min.]

and setting $\gamma_i = z_i^{-1/2}|x_{*,i}|$.

Thus, the iterative rewt. ℓ_1 proc.:

Init: $z_i = 1 \ \forall i$

Repeat

$\quad \gamma \leftarrow \arg\min_\gamma \ L_z(\gamma) \triangleq z^T\gamma + y^T\Sigma_y^{-1}y$

$\quad z \leftarrow \nabla_\gamma \log\det\Sigma_y$

Until convergence of γ to γ_*

Output: $x_* = E\{x \mid y; \Gamma_*\} = \Gamma_* A^T \Sigma_{y*}^{-1} y$.

Iterative reweighted ℓ_2 approach:

$$\min_{\gamma \geq 0} \ L(\gamma) = \log\det\Sigma_y + y^T\Sigma_y^{-1}y$$
$$\Sigma_y = \sigma^2 I + A\Gamma A^T, \quad \Gamma = \operatorname{diag}(\gamma).$$

After we obtain γ_*, the soln. is
$$x_{SBL} = E\{x \mid y; \gamma_*\} = \Gamma_* A^T \Sigma_{y*}^{-1} y.$$

$$y^T\Sigma_y^{-1}y = \min_x \frac{\|y - Ax\|_2^2}{\sigma^2} + x^T\Gamma^{-1}x$$

Substituting into the SBL cost fn. and exchanging the order of min.,
$$\min_x \frac{\|y - Ax\|_2^2}{\sigma^2} + g_{SBL}(x)$$

where $g_{SBL}(x) \triangleq \min_{\gamma \geq 0} x^T\Gamma^{-1}x + \log\det\Sigma_y$

Can s.t. $g_{SBL}(x)$ (Appendix of David Wipf, Iterative ℓ_1-ℓ_2 methods for SBL) is nondecreasing and concave in $|x|$, and $|x|^2$.

$|x|^2 = [|x_1|^2, |x_2|^2, \ldots, |x_N|^2]^T$.

$$\Rightarrow g\ (x) \leq x^T\Gamma^{-1}x + \log\det(\sigma^2 I + A\Gamma A^T)$$

$\rightarrow f_{SBL}$

$$= x^T\Gamma^{-1}x + \log\det\Gamma + \underline{\log\det(\sigma^{-2}A^TA + \Gamma^{-1})} + m\log\sigma^2$$

$$\leq \underline{x^T\Gamma^{-1}x + \log\det\Gamma + z^T\gamma^{-1} - h^*(z)} + m\log\sigma^2$$

where $h^*(z)$ = concave conjugate of $h(z)$

$$h(z) = \log\det(\sigma^{-2}A^TA + \mathrm{diag}(z))$$

$$h^*(z) = \min_{a \geq 0} z^Ta - \log\det(\sigma^{-2}A^TA + \mathrm{diag}(a))$$

Thus, we perform alt. min. using

$$\min_{\gamma, x, z \geq 0} \|y - Ax\|_2^2 + \sigma^2\left[-h^*(z) + \sum_{i=1}^{N} \frac{x_i^2 + z_i^2}{\gamma_i} + \log\gamma_i\right]$$

$$x^{(k+1)} = \tilde{W}^{(k)}A^T(\sigma^2 I + A\tilde{W}^{(k)}A^T)^{-1}y,$$

where $\tilde{W}^{(k)} = \mathrm{diag}(\gamma_i^{(k)})$

The update for z_i comes from the defn. of the concave conjugate fn.,

$$\underline{z_i^{(k+1)}} = \frac{\partial}{\partial a_i}\log\det(\sigma^{-2}A^TA + \mathrm{diag}(a))$$

$$= \left[(\sigma^{-2}A^TA + \mathrm{diag}(\gamma^{(k)})^{-1})^{-1}\right]_{ii}$$

$$= \gamma_i^{(k)} - {\gamma_i^{(k)}}^2 a_i^T(\sigma^2 I + A\Gamma^{(k)}A^T)^{-1}a_i$$

$$i = 1, 2, \ldots, N.$$

$$\underline{\gamma_i^{(k+1)}} = \underline{(x_i^{(k)})}^2 + (z_i^{(k)})^2.$$

We iterate the three steps above till convergence.

Extensions: Non-neg. sparse recovery:

Just include the constraint $x \geq 0$ in ⊛.

Group sparsity:

$$X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_L], \quad \underline{x}_i \in \mathbb{R}^N$$

sparse vectors with common support.

Let $x_{i\cdot} \triangleq i^{th}$ row of X.

$$d(X) = \sum_{i=1}^{N} \mathbb{1}\{\|x_{i\cdot}\| > 0\}$$

$$\min_X d(X) \quad \text{s.t.} \quad Y = AX + W$$

$$[\underline{y}_1 \cdots \underline{y}_L] \in \mathbb{R}^{m \times L}$$

The above algos can be extended easily:

Replace $|x_i^{(k)}|$ with l_2 norm of the row $(x_{i\cdot}^{(k)})_{1 \times L}$.
