Gradient Computation

ss2571

May 2020

This document contains the gradient formula for the linear and nonlinear lasso problem.

1   Gradient computation

Define two points in spherical coordinates as follows:

$$\boldsymbol{Q} = (1, \theta, \psi), \ \boldsymbol{P_i} = (r_i, \theta_i, \psi_i). \tag{1}$$

The normalized forward solution reads:

$$\hat{\Phi}(\mu)(\boldsymbol{Q}) = \sum_{i=1}^{N} \frac{I_i}{G_{\boldsymbol{P_i}}} \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}), \ \boldsymbol{Q} \in \partial\Omega, \ where, \tag{2}$$

$$G_{\boldsymbol{P_i}} = \sqrt{\int_{\partial\Omega} |F(\boldsymbol{Q}, \boldsymbol{P_i})|^2 d\boldsymbol{Q}} = \sqrt{\int_{\partial\Omega} \phi_{\boldsymbol{P_i}}^2(\boldsymbol{Q}) d\boldsymbol{Q}} \tag{3}$$

where,

$$\begin{aligned}
\phi_{\boldsymbol{P_i}}(\boldsymbol{Q}) &= \frac{2}{||\boldsymbol{Q} - \boldsymbol{P_i}||_2} - \log(1 - \langle \boldsymbol{P_i}, \ \boldsymbol{Q} \rangle + ||\boldsymbol{Q} - \boldsymbol{P_i}||_2) \\
&= \frac{2}{\sqrt{1 + r_i^2 - 2r_i \cos\gamma_i}} - \log(1 - r_i \cos\gamma_i + \sqrt{1 + r_i^2 - 2r_i \cos\gamma_i}), \ where,
\end{aligned} \tag{4}$$

$$\cos\gamma_i = \cos\theta\cos\theta_i + \sin\theta\sin\theta_i\cos(\psi - \psi_i), \ \langle \boldsymbol{P_i}, \boldsymbol{Q} \rangle = r_i\cos\gamma_i, \ ||\boldsymbol{Q} - \boldsymbol{P_i}||_2 = \sqrt{1 + r_i^2 - 2r_i\cos\gamma_i}.$$

Define the source parameter vector as follows:

$$\boldsymbol{V} = \{\overbrace{I_1, I_2, \cdots}^{N}, \overbrace{r_1, r_2, \cdots}^{N}, \overbrace{\theta_1, \theta_2, \cdots}^{N}, \overbrace{\psi_1, \psi_2, \cdots}^{N}\}, \tag{5}$$

$$\boldsymbol{I_{all}} = \{I_1, I_2, \cdots, I_N\}. \tag{6}$$

And in this report, we use the symbol D^i to represent the partial derivative with respect to the *ith* component of the unknown parameter vector.

1.1   Linear lasso problem

For the linear lasso problem, locations of the source (i.e., $\boldsymbol{P_i}$) are given, we define the forward estimation as follows:

$$Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) = \sum_{i=1}^{N} \frac{I_i}{G_{\boldsymbol{P_i}}} \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}), \ \boldsymbol{Q} \in \partial\Omega, \tag{7}$$

The objective function reads:

$$Ojb = \frac{1}{2} ||Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})||_2^2 + \lambda||\boldsymbol{I_{all}}||_1, \tag{8}$$

$$= \int_{\partial\Omega} [Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q} + \lambda \sum_i |I_i|. \tag{9}$$

For convenience, we introduce the following two terms:

$$f_{des}(\boldsymbol{I_{all}}) := \frac{1}{2}||Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})||_2^2 = \frac{1}{2}\int_{\partial\Omega}[Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q}, \quad (10)$$

$$f_{reg} := \lambda||\boldsymbol{I_{all}}||_1. \quad (11)$$

1.2 Subgradient of the regularization term

$$D^i f_{reg} = \lambda\, sign(I_i). \quad (12)$$

FBS

For the FBS solver, only the gradient of the discrepancy term is required, and locations of the source are given. Therefore we only compute:

$$Grad_i = D^i f_{des}(\boldsymbol{Q}) = \int_{\partial\Omega}(Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q}))D^i Est(\boldsymbol{I_{all}}, \boldsymbol{Q})d\boldsymbol{Q}, \quad (13)$$

$$= \int_{\partial\Omega}(Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q}))\frac{1}{G_{\boldsymbol{P_i}}}\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})d\boldsymbol{Q},$$

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

In the minimization iteration, source location $\boldsymbol{P_i}$ is provided by the solver in each search step.

Quasi Newton

For quasi newton solver, gradient of the whole objective function needs to be computed.

$$PseudoGrad_i = \int_{\partial\Omega}(Est(\boldsymbol{I_{all}}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q}))\frac{1}{G_{\boldsymbol{P_i}}}\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})d\boldsymbol{Q} + \lambda\, sign(I_i). \quad (14)$$

The Lipschitz constant of the discrepancy gradient

Proposition 1.1. *Let $\boldsymbol{I_{all,1}}, \boldsymbol{I_{all,2}}$ be two intensity solution vectors and $I_{i,1}, I_{i,2}$ be their components, and let:*

$$||\boldsymbol{x}|| := \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2}, \text{ for any } \boldsymbol{x} \in R^N.$$

Define a vector as follows:

$$\boldsymbol{A}(\boldsymbol{Q}) := \{D^1 Est(\boldsymbol{I_{all}}, \boldsymbol{Q}), D^2 Est(\boldsymbol{I_{all}}, \boldsymbol{Q}), \cdots, D^N Est(\boldsymbol{I_{all}}, \boldsymbol{Q})\}$$

$$= \{\frac{1}{G_{\boldsymbol{P_1}}}\phi_{\boldsymbol{P_1}}(\boldsymbol{Q}), \frac{1}{G_{\boldsymbol{P_2}}}\phi_{\boldsymbol{P_2}}(\boldsymbol{Q}), \cdots, \frac{1}{G_{\boldsymbol{P_N}}}\phi_{\boldsymbol{P_N}}(\boldsymbol{Q})\}.$$

, then the gradient $\nabla f_{des}(\boldsymbol{I_{all}})$ has a Lipschitz constant as follows:

$$L = |\sum_i \int_{\partial\Omega}||A(\boldsymbol{Q})||A_i(\boldsymbol{Q})d\boldsymbol{Q}|. \quad (15)$$

Proof. Notice that:

$$||\nabla f_{des}(\boldsymbol{I_{all,1}}) - \nabla f_{des}(\boldsymbol{I_{all,2}})|| = ||\int_{\partial\Omega}[Est(\boldsymbol{I_{all,1}}, \boldsymbol{Q}) - Est(\boldsymbol{I_{all,2}}, \boldsymbol{Q})] \cdot \nabla Est(\boldsymbol{I_{all}}, \boldsymbol{Q})d\boldsymbol{Q}||$$

$$= ||\int_{\partial\Omega}\langle\boldsymbol{A}(\boldsymbol{Q}), \boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}}\rangle\boldsymbol{A}(\boldsymbol{Q})d\boldsymbol{Q}||, \quad (16)$$

Using the Cauchy Schwarz inequality to the vector l2 norm, we get:

$$\int_{\partial\Omega}\langle\boldsymbol{A}(\boldsymbol{Q}), \boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}}\rangle\boldsymbol{A}(\boldsymbol{Q})d\boldsymbol{Q} \le \int_{\partial\Omega}||\boldsymbol{A}(\boldsymbol{Q})||\,||\boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}}||\boldsymbol{A}(\boldsymbol{Q})d\boldsymbol{Q}. \quad (17)$$

Return to Equation (16), we derive the following:

$$||\nabla f_{des}(\boldsymbol{I_{all,1}}) - \nabla f_{des}(\boldsymbol{I_{all,2}})|| = ||\int_{\partial\Omega} \langle \boldsymbol{A}(\boldsymbol{Q}), \boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}} \rangle \boldsymbol{A}(\boldsymbol{Q})d\boldsymbol{Q}|| \tag{18}$$

$$\leq \sqrt{\sum_i [\ ||\boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}}|| \int_{\partial\Omega} ||A(\boldsymbol{Q})\ ||A_i(\boldsymbol{Q})d\boldsymbol{Q}\ ]^2} \tag{19}$$

$$= |\ \sum_i \int_{\partial\Omega} ||A(\boldsymbol{Q})\ ||A_i(\boldsymbol{Q})d\boldsymbol{Q}\ |\ ||\boldsymbol{I_{all,1}} - \boldsymbol{I_{all,2}}|| \tag{20}$$

therefore, we obtain an estimation of the Lipschitz constant as follows,

$$L = |\ \sum_i \int_{\partial\Omega} ||A(\boldsymbol{Q})\ ||A_i(\boldsymbol{Q})d\boldsymbol{Q}\ |. \tag{21}$$

$\square$

1.3   Nonlinear lasso problem

For the nonlinear lasso problem, both locations and intensities are unknown, the forward estimation reads:

$$Est(\boldsymbol{V}, \boldsymbol{Q}) = \sum_{i=1}^N \frac{I_i}{G_{\boldsymbol{P_i}}} \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}),\ \boldsymbol{Q} \in \partial\Omega, \tag{22}$$

The objective function reads:

$$Ojb = \frac{1}{2}||Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})||_2^2 + \lambda||I_{all}||_1, \tag{23}$$

$$= \int_{\partial\Omega} [Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q} + \lambda \sum_i |I_i|. \tag{24}$$

The pseudo gradient of the nonlinear lasso problem consists of two component: the gradient of the discrepancy term and the subgradient of the l1 regularizatoin term.

The discrepancy gradient

In this section we compute the gradient term ∇f_{des}.

$$For\ i = 1 \rightarrow N,$$

$$Grad_i = D^i f_{des}(\boldsymbol{Q}) = D^i f_{des}(\boldsymbol{Q}) = \int_{\partial\Omega} (Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})) \frac{1}{G_{\boldsymbol{P_i}}} \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}) d\boldsymbol{Q}, \tag{25}$$

$$for\ i = N + 1 \rightarrow 4N,$$

$$Grad_i = D^i f_{des}(\boldsymbol{Q}) = D^i f_{des}(\boldsymbol{Q}) = \int_{\partial\Omega} (Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})) D^i Est(\boldsymbol{V}, \boldsymbol{Q}) d\boldsymbol{Q}, \tag{26}$$

$$D^i Est(\boldsymbol{V}, \boldsymbol{Q}) = I_i \frac{G_{\boldsymbol{P_i}} D^i \phi_{\boldsymbol{P_i}} - \phi_{\boldsymbol{P_i}} D^i G_{\boldsymbol{P_i}}}{G_{\boldsymbol{P_i}}^2}, \tag{27}$$

Recall that:

$$\phi_{\boldsymbol{P_i}}(\boldsymbol{Q}) = \frac{2}{||\boldsymbol{Q} - \boldsymbol{P_i}||_2} - \log(1 - \langle \boldsymbol{P_i},\ \boldsymbol{Q} \rangle + ||\boldsymbol{Q} - \boldsymbol{P_i}||_2),$$

$$G_{\boldsymbol{P_i}} = \sqrt{\int_{\partial\Omega} |F(\boldsymbol{Q}, \boldsymbol{P_i})|^2 d\boldsymbol{Q}} = \sqrt{\int_{\partial\Omega} \phi_{\boldsymbol{P_i}}^2(\boldsymbol{Q}) d\boldsymbol{Q}}.$$

Thus,

$$D^i \phi_{\boldsymbol{P_i}} = -\frac{1}{||\boldsymbol{Q} - \boldsymbol{P_i}||_2^3} D^i ||\boldsymbol{Q} - \boldsymbol{P_i}||_2 - \frac{1}{1 - \langle \boldsymbol{P_i},\ \boldsymbol{Q} \rangle + ||\boldsymbol{Q} - \boldsymbol{P_i}||_2} D^i(-\langle \boldsymbol{P_i},\ \boldsymbol{Q} \rangle + ||\boldsymbol{Q} - \boldsymbol{P_i}||_2). \tag{28}$$

3

$$D^i G_{\boldsymbol{P_i}} = \frac{\int_{\partial\Omega} \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}) D^i \phi_{\boldsymbol{P_i}}(\boldsymbol{Q}) d\boldsymbol{Q}}{G_{\boldsymbol{P_i}}}. \tag{29}$$

Next we compute the following terms:

$$D^i \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2, \quad D^i \langle \boldsymbol{Q}, \boldsymbol{P_i} \rangle, \ where,$$

$$\langle \boldsymbol{P_i}, \boldsymbol{Q} \rangle = r_i \cos \gamma_i, \ \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2 = \sqrt{1 + r_i^2 - 2r_i \cos \gamma_i},$$

$$\cos \gamma_i = \cos \theta \cos \theta_i + \sin \theta \sin \theta_i \cos(\psi - \psi_i).$$

For $i = N + 1 \to 2N$:

$$D^i \langle \boldsymbol{Q}, \boldsymbol{P_i} \rangle = \cos \gamma_i, \tag{30}$$

$$D^i \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2 = \frac{\partial \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}{\partial r_i} = \frac{r_i - \cos \gamma_i}{\|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}, \tag{31}$$

for $i = 2N + 1 \to 3N$:

$$D^i \langle \boldsymbol{Q}, \boldsymbol{P_i} \rangle = r_i(-\cos \theta \sin \theta_i + \sin \theta \cos \theta_i \cos(\psi - \psi_i)), \tag{32}$$

$$D^i \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2 = \frac{\partial \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}{\partial \theta_i} = \frac{-r_i(-\cos \theta \sin \theta_i + \sin \theta \cos \theta_i \cos(\psi - \psi_i))}{\|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}, \tag{33}$$

for $i = 3N + 1 \to 4N$:

$$D^i \langle \boldsymbol{Q}, \boldsymbol{P_i} \rangle = r_i(\cos \theta \cos \theta_i - \sin \theta \sin \theta_i \sin(\psi - \psi_i), \tag{34}$$

$$D^i \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2 = \frac{\partial \|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}{\partial \psi_i} = \frac{-r_i(-\sin \theta \sin \theta_i \sin(\psi - \psi_i))}{\|\boldsymbol{Q} - \boldsymbol{P_i}\|_2}. \tag{35}$$

Pseudo Gradient

Whenever the gradient of the whole objective function is required, we compute the discrepancy gradient and add the subgradient of the regularization term to it, i.e.,

$$PsudoGrad_i = \begin{cases} Grad_i + \lambda * sign(I_i), \ i = 1 \to N, \\ Grad_i, \ i > N. \end{cases} \tag{36}$$