Gradient Computation

ss2571

May 2020

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

1 Setup

Define two points in spherical coordinates:

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

The normalized forward solution reads:

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

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

where,

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

Define the source parameter vector as follows:

$$\boldsymbol{V} = \{\overbrace{I_1, I_2, \cdots,}^{N} \overbrace{\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_N}^{3N}\}, \tag{5}$$

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

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.

2 Linear lasso problem

For the linear lasso problem, locations of the source (for example, $\boldsymbol{P_i}$) are given, we define the estimation function as follows:

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

Linear forward operator and its adjoint operator

Define a vector function $\boldsymbol{A(Q)} \in R^N$ such that,

$$\boldsymbol{A(Q)} = \{A_1(\boldsymbol{Q}), A_2(\boldsymbol{Q}), \cdots, A_N(\boldsymbol{Q})\},\ and, \tag{8}$$

$$A_i(\boldsymbol{Q}) := \frac{\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})}{G_{\boldsymbol{P_i}}} \in L^2(\partial\Omega). \tag{9}$$

We define the linear forward operator as follows:

$$\Phi : R^N \to L^2(\partial\Omega),\ s.t.,$$

$$\Phi(\boldsymbol{I}) = \langle \boldsymbol{A(Q)}, \boldsymbol{I} \rangle = \sum_{i=1}^{N} I_i A_i(\boldsymbol{Q}). \tag{10}$$

Then, the estimation function becomes:

$$Est(\boldsymbol{I}, \boldsymbol{Q}) = \Phi(\boldsymbol{I}) \tag{11}$$

Next we define the adjoint of Φ, denote it as Φ^*:

$$\Phi^* : L^2(\partial\Omega) \to R^N,\ s.t., \tag{12}$$

$$\Phi^*(p(\boldsymbol{Q})) = \int_{\partial\Omega} \boldsymbol{A(Q)} p(\boldsymbol{Q}) d\boldsymbol{Q},\ p(\boldsymbol{Q}) \in L^2(\partial\Omega). \tag{13}$$

The composition map $\Phi^*\Phi$ derives from the following steps, by definition,

$$\Phi^*\Phi : R^N \to R^N. \tag{14}$$

Take arbitrary vector $a \in R^N$, and apply the map, we get:

$$\Phi^*\Phi(\boldsymbol{a}) = \int_{\partial\Omega} \boldsymbol{A(Q)} \sum_{i=1}^{N} A_i(\boldsymbol{Q}) a_i d\boldsymbol{Q} \in R^N, \tag{15}$$

take the kth entry of $\Phi^*\Phi(\boldsymbol{a})$, we derive:

$$[\Phi^*\Phi(\boldsymbol{a})]_k = \int_{\partial\Omega} A_k(\boldsymbol{Q}) A_i(\boldsymbol{Q}) a_i d\boldsymbol{Q} = \sum_{i=1}^{N} (\int_{\partial\Omega} A_k(\boldsymbol{Q}) A_i(\boldsymbol{Q}) d\boldsymbol{Q}) a_i = \sum_{i=1}^{N} M_{k,i} a_i \tag{16}$$

, from above we derive that,

$$\mathcal{M}_{i,j} := (\Phi^*\Phi)_{i,j} = \int_{\partial\Omega} A_i(\boldsymbol{Q}) A_j(\boldsymbol{Q}) d\boldsymbol{Q}. \tag{17}$$

Therefore,

$$\Phi^*\Phi = \mathcal{M} \in R^{N,N}. \tag{18}$$

Objective function

The objective function reads:

$$Obj(\boldsymbol{I}) = \frac{1}{2} ||\Phi(\boldsymbol{I}) - \phi^d(\boldsymbol{Q})||^2_{\partial\Omega} + \lambda ||\boldsymbol{I}||_1, \tag{19}$$

$$= \int_{\partial\Omega} [\Phi(\boldsymbol{I}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q} + \lambda \sum_i |I_i|. \tag{20}$$

For convenience, we introduce the following two terms:

$$f_{des}(\boldsymbol{I}) := \frac{1}{2} ||\Phi(\boldsymbol{I}) - \phi^d(\boldsymbol{Q})||^2_{\partial\Omega} = \frac{1}{2} \int_{\partial\Omega} [\Phi(\boldsymbol{I}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q}, \tag{21}$$

$$f_{reg}(\boldsymbol{I}) := \lambda ||\boldsymbol{I}||_1. \tag{22}$$

2.1 The gradient and hessian of f_{des}

Through a direct calculation, we find that,

$$\nabla f_{des}(\boldsymbol{I}) = \Phi^*\Phi(\boldsymbol{I}) - \Phi^*(\phi^d) = \mathcal{M}\boldsymbol{I} - \boldsymbol{b},\ where,\ \boldsymbol{b} = \Phi^*(\phi^d). \tag{23}$$

Follows from above, we find the hessian of f_{des},

$$\mathcal{H}(f_{des}(\boldsymbol{I})) = \mathcal{M}. \tag{24}$$

2.1.1 The Lipschitz constant of ∇f_{des}

Proposition 2.1. *The gradient $\nabla f_{des}(\boldsymbol{I})$ has a Lipschitz constant as follows:*

$$L = ||\boldsymbol{\Phi^*\Phi}||_2,\ where\ ||\cdot||_2\ is\ the\ matrix\ l2\ norm. \tag{25}$$

Proof. Let $\boldsymbol{I}_a, \boldsymbol{I}_b \in R^N$ be two arbitrary vectors, then,

$$||\nabla f_{des}(\boldsymbol{I}_a) - \nabla f_{des}(\boldsymbol{I}_b)||_2 = ||\mathcal{M}(\boldsymbol{I}_a - \boldsymbol{I}_b)||_2 \leq ||\mathcal{M}||_2||\boldsymbol{I}_a - \boldsymbol{I}_b||_2. \tag{26}$$

□

2.2 The subgradient of f_{reg}

Denote the subgradient of f_{reg} as ∂f_{reg}, we have that,

$$\partial f_{reg}(\boldsymbol{I}) = \lambda\ sign(\boldsymbol{I}). \tag{27}$$

2.3 Gradient calculation for linear solvers

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:

$$\nabla f_{des}(\boldsymbol{I}) = \mathcal{M}\boldsymbol{I} - \Phi^*(\phi^d(\boldsymbol{Q})),\ where, \tag{28}$$

$$\mathcal{M}_{i,j} = \int_{\partial\Omega} A_i(\boldsymbol{Q})A_j(\boldsymbol{Q})d\boldsymbol{Q},\ and,\ \Phi^*(\phi^d(\boldsymbol{Q})) = \int_{\partial\Omega} \boldsymbol{A(Q)}\phi^d(\boldsymbol{Q})d\boldsymbol{Q}. \tag{29}$$

Quasi Newton

For quasi newton solver, we compute the gradient of the entire objective function, i.e.,

$$\nabla_{pseudo}Obj(\boldsymbol{I}) = \nabla f_{des} + \partial f_{reg}. \tag{30}$$

Proof. □

3 Nonlinear lasso problem

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

$$Est(\boldsymbol{V}, \boldsymbol{Q}) = \sum_{i=1}^{N} \frac{I_i}{G_{\boldsymbol{P_i}}}\phi_{\boldsymbol{P_i}}(\boldsymbol{Q}),\ \boldsymbol{Q} \in \partial\Omega, \tag{31}$$

The objective function reads:

$$Ojb = \frac{1}{2}||Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})||^2_{\partial\Omega} + \lambda||I_{all}||_1, \tag{32}$$

$$= \frac{1}{2}\int_{\partial\Omega}[Est(\boldsymbol{V}, \boldsymbol{Q}) - \phi^d(\boldsymbol{Q})]^2 d\boldsymbol{Q} + \lambda\sum_i |I_i|. \tag{33}$$

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}.

$$
\begin{aligned}
&For\ i=1\rightarrow N,\\
&Grad_i = D^i f_{des}(\boldsymbol{Q}) = D^i f_{des}(\boldsymbol{Q}) = \int_{\partial\Omega}(Est(\boldsymbol{V},\boldsymbol{Q})-\phi^d(\boldsymbol{Q}))\frac{1}{G_{\boldsymbol{P_i}}}\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})d\boldsymbol{Q}, &(34)\\
&for\ i=N+1\rightarrow 4N,\\
&Grad_i = D^i f_{des}(\boldsymbol{Q}) = D^i f_{des}(\boldsymbol{Q}) = \int_{\partial\Omega}(Est(\boldsymbol{V},\boldsymbol{Q})-\phi^d(\boldsymbol{Q}))D^i Est(\boldsymbol{V},\boldsymbol{Q})d\boldsymbol{Q}, &(35)\\
&D^i Est(\boldsymbol{V},\boldsymbol{Q}) = I_i\frac{G_{\boldsymbol{P_i}}D^i\phi_{\boldsymbol{P_i}}-\phi_{\boldsymbol{P_i}}D^iG_{\boldsymbol{P_i}}}{G^2_{\boldsymbol{P_i}}}, &(36)
\end{aligned}
$$

Recall that:

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

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

Thus,

$$D^i\phi_{\boldsymbol{P_i}} = -\frac{1}{||\boldsymbol{Q}-\boldsymbol{P_i}||_2^3}D^i||\boldsymbol{Q}-\boldsymbol{P_i}||_2 - \frac{1}{1-\langle \boldsymbol{P}_i,\ \boldsymbol{Q}\rangle + ||\boldsymbol{Q}-\boldsymbol{P_i}||_2}D^i(-\langle \boldsymbol{P}_i,\ \boldsymbol{Q}\rangle + ||\boldsymbol{Q}-\boldsymbol{P_i}||_2). \tag{37}$$

$$D^iG_{\boldsymbol{P_i}} = \frac{\int_{\partial\Omega}\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})D^i\phi_{\boldsymbol{P_i}}(\boldsymbol{Q})d\boldsymbol{Q}}{G_{\boldsymbol{P_i}}}. \tag{38}$$

Next we compute the following terms:

$$
\begin{aligned}
&D^i||\boldsymbol{Q}-\boldsymbol{P_i}||_2,\quad D^i\langle\boldsymbol{Q},\boldsymbol{P_i}\rangle,\ where,\\
&\langle\boldsymbol{P}_i,\boldsymbol{Q}\rangle = r_i\cos\gamma_i,\ ||\boldsymbol{Q}-\boldsymbol{P_i}||_2 = \sqrt{1+r_i^2-2r_i\cos\gamma_i},\\
&\cos\gamma_i = \cos\theta\cos\theta_i + \sin\theta\sin\theta_i\cos(\psi-\psi_i).
\end{aligned}
$$

$$
\begin{aligned}
&\text{For } i=N+1:3:4N-2,\\
&D^i\langle\boldsymbol{Q},\boldsymbol{P_i}\rangle = \cos\gamma_i, &(39)\\
&D^i||\boldsymbol{Q}-\boldsymbol{P_i}||_2 = \frac{\partial||\boldsymbol{Q}-\boldsymbol{P_i}||_2}{\partial r_i} = \frac{r_i-\cos\gamma_i}{||\boldsymbol{Q}-\boldsymbol{P_i}||_2}, &(40)\\
&\text{for } i=N+2:3:4N-1,\\
&D^i\langle\boldsymbol{Q},\boldsymbol{P_i}\rangle = r_i(-\cos\theta\sin\theta_i+\sin\theta\cos\theta_i\cos(\psi-\psi_i)), &(41)\\
&D^i||\boldsymbol{Q}-\boldsymbol{P_i}||_2 = \frac{\partial||\boldsymbol{Q}-\boldsymbol{P_i}||_2}{\partial\theta_i} = \frac{-r_i(-\cos\theta\sin\theta_i+\sin\theta\cos\theta_i\cos(\psi-\psi_i))}{||\boldsymbol{Q}-\boldsymbol{P_i}||_2}, &(42)\\
&\text{for } i=N+3:3:4N,\\
&D^i\langle\boldsymbol{Q},\boldsymbol{P_i}\rangle = r_i(\cos\theta\cos\theta_i-\sin\theta\sin\theta_i\sin(\psi-\psi_i), &(43)\\
&D^i||\boldsymbol{Q}-\boldsymbol{P_i}||_2 = \frac{\partial||\boldsymbol{Q}-\boldsymbol{P_i}||_2}{\partial\psi_i} = \frac{-r_i(-\sin\theta\sin\theta_i\sin(\psi_i-\psi))}{||\boldsymbol{Q}-\boldsymbol{P_i}||_2}. &(44)
\end{aligned}
$$

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.,

$$PsudoGrad_i = \begin{cases} Grad_i + \lambda * sign(I_i),\ i=1\rightarrow N,\\ Grad_i,\ i>N.\end{cases} \tag{45}$$