2022-01-14

Victor

January 18, 2022

Contents

1 Réunion Arthur/Olivier **1**
1.1 Autoencoders with Invertible maps 1

2 Arthur: GN **1**
2.1 3DVar: Cost function, Gradient and Hessian 1
2.2 Gauss-Newton method 2
2.3 AE formulation. 2

1 Réunion Arthur/Olivier

1.1 Autoencoders with Invertible maps

$$\begin{aligned} x &= \phi^{-1} \circ \phi(x) && (1) \\ &= \phi^{-1}\left(U_m U_m^T \phi(x) + U_\perp U_\perp^T \phi(x)\right) && (2) \end{aligned}$$

Where $U = [U_m \quad U_\perp]$ is unitary

$$\begin{aligned} E(x) &= U_m^T \phi(x) && (3) \\ D(z) &= \phi^{-1}(U_m z + U_\perp z_0) && (4) \end{aligned}$$

Construction of ϕ ? ResNet, RevNet ?

2 Arthur: GN

2.1 3DVar: Cost function, Gradient and Hessian

Revoir [Gratton et al., 2007]

$$J(x) = \frac{1}{2}\|(\mathcal{H} \circ \mathcal{M})(x) - y\|_R^2 + \frac{1}{2}\|x - x_b\|_B^2 \tag{5}$$
$$= \frac{1}{2}\|G(x) - y\|_R^2 + \frac{1}{2}\|x - x_b\|_B^2 \tag{6}$$

with $G = \mathcal{H} \circ \mathcal{M}$

The gradient of J is then

$$\nabla J(x) = \nabla G(x)^T R^{-1} (G(x) - y) + B^{-1}(x - x_b) \tag{7}$$

and the Hessian is

$$\nabla^2 G(x) = \left(\nabla G(x)^T R^{-1} \nabla G(x) + B^{-1}\right) + Q(x) \tag{8}$$

2.2 Gauss-Newton method

Per, Approximate GN methods for non-linear least-square problems The Gauss-Newton method relies on the knowledge of the GN-Hessian

$$\nabla G(x)^T R^{-1} \nabla G(x) + R^{-1} \tag{9}$$

2.3 AE formulation.

Let $G = D \circ E \Rightarrow G(x) = D(E(x))$

$$\nabla G(x) = \nabla D(E(x)) \nabla E(x) \tag{10}$$

So

$$\nabla G(x)^T R^{-1} \nabla G(x) = \nabla E(x)^T \nabla D(E(x))^T R^{-1} \nabla D(E(x)) \nabla E(x) \tag{11}$$
$$= \nabla E(x)^T \left(\nabla D(E(x))^T R^{-1} \nabla D(E(x))\right) \nabla E(x) \tag{12}$$

Let $x \in \mathbb{R}^n$, $G(x) \in \mathbb{R}^m$, and $E(x) \in \mathbb{R}^p$

Where $n > m > p$

$$\nabla E(x) : \mathbb{R}^n \to \mathbb{R}^{p \times n} \tag{13}$$
$$\nabla D(x) : \mathbb{R}^p \to \mathbb{R}^{m \times p} \tag{14}$$
$$\nabla D(E(x)) : \mathbb{R}^n \to \mathbb{R}^{m \times p} \tag{15}$$

So

$$\nabla G(x)^T R^{-1} \nabla G(x) = \underbrace{\nabla E(x)^T}_{\in \mathbb{R}^{n \times p}} \underbrace{\left(\nabla D(E(x))^T R^{-1} \nabla D(E(x))\right)}_{\in \mathbb{R}^{p \times p}} \underbrace{\nabla E(x)}_{\in \mathbb{R}^{p \ \times n}} \tag{16}$$

Construct:

$$\nabla G(x^k) R^{-1} \nabla G(x^k) \tag{17}$$

References

[Gratton et al., 2007] Gratton, S., Lawless, A. S., and Nichols, N. K. (2007). Approximate Gauss–Newton Methods for Nonlinear Least Squares Problems. *SIAM Journal on Optimization*, 18(1):106–132.