量子几何张量 vs Fisher 信息矩阵

从 KL 散度到 Fubini-Study 距离，从自然梯度到随机重构的自洽推导

摘要

经典信息几何中，KL 散度在小参数扰动下的二阶展开诱导出 Fisher-Rao 度量，其度量张量即 Fisher 信息矩阵（FIM）；对应的”最速下降”给出自然梯度法。量子纯态的物理态是射线（ray），存在全局相位与整体归一化的规范冗余，因而不能直接把”对输出的欧氏内积/协方差”当作度量；正确的距离应定义在射影 Hilbert 空间上，即 Fubini-Study（FS）距离。FS 距离的二阶展开给出量子几何张量（QGT）：其实部是 FS 黎曼度量，虚部是 Berry 曲率。进一步地，在以 FS 范数约束的最速下降问题中，参数更新满足线性方程 $S\delta\theta = -\eta g$，这正是变分蒙特卡洛与神经网络量子态训练中常用的随机重构（SR）/量子自然梯度（QNG）更新。本文从经典部分开始，给出 KL 一阶项为零、二阶项等于 score 协方差的完整证明，并在清晰区分”概率输出”和”复振幅输出”的基础上，逐步引入 FS 距离、QGT 的规范不变定义及其与 SR 的几何推导。文末用两张对照表总结经典与量子概念的对应关系，以及常见优化算法与其隐含的度量/流形解释（包括 Adam/RMSProp 等对角近似）。

目录

1 经典信息几何：KL 散度的局部展开与 Fisher-Rao 度量

1.1 从 KL 散度到 Fisher 信息矩阵

设 $p_{\boldsymbol{\theta}}(x)$ 是由一组参数 $\boldsymbol{\theta}=(\theta^1,\dots,\theta^d)$ 参数化的概率分布（离散与连续情形统一记作 $\sum_x$）。两分布之间的差异由 Kullback-Leibler (KL) 散度刻画：

$$D_{\mathrm{KL}}\left(p_{\boldsymbol{\theta}}\|p_{\boldsymbol{\theta}'}\right):=\sum_x p_{\boldsymbol{\theta}}(x)\ln\frac{p_{\boldsymbol{\theta}}(x)}{p_{\boldsymbol{\theta}'}(x)}=\sum_x p_{\boldsymbol{\theta}}(x)\ln p_{\boldsymbol{\theta}}(x)-\sum_x p_{\boldsymbol{\theta}}(x)\ln p_{\boldsymbol{\theta}'}(x). \tag{1}$$

注意 KL 散度不对称，所以严格来说它不是一种“距离”。经典信息几何把分布族 $\{p_{\boldsymbol{\theta}}\}$ 看作流形上的点集，我们仍希望在该流形上定义“距离”。一个自然思路是：当两分布非常接近时，用某种距离度量来近似 KL 散度。

考虑参数的微小变动 $\boldsymbol{\theta}'=\boldsymbol{\theta}+\delta\boldsymbol{\theta}$，对 KL 散度做泰勒展开（即展开 $\ln p_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}(x)$）：

$$\begin{aligned}
D_{\mathrm{KL}}\left(p_{\boldsymbol{\theta}}\|p_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}\right)&=\sum_x p_{\boldsymbol{\theta}}(x)\ln p_{\boldsymbol{\theta}}(x)-\sum_x p_{\boldsymbol{\theta}}(x)\ln p_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}(x)\\
&=\sum_x p_{\boldsymbol{\theta}}(x)\underbrace{\left[\ln p_{\boldsymbol{\theta}}(x)-\ln p_{\boldsymbol{\theta}}(x)\right]}_{=0} && \text{（零阶项）}\\
&-\delta\theta^i\underbrace{\sum_x p_{\boldsymbol{\theta}}(x)\partial_i\ln p_{\boldsymbol{\theta}}(x)}_{=0} && \text{（一阶项）}\\
&-\frac{1}{2}\delta\theta^i\delta\theta^j\underbrace{\sum_x p_{\boldsymbol{\theta}}(x)\partial_i\partial_j\ln p_{\boldsymbol{\theta}}(x)}_{:=-I_{ij}(\boldsymbol{\theta})} && \text{（二阶项）}\\
&+\mathrm{O}(\|\delta\boldsymbol{\theta}\|^3)
\end{aligned}$$

其中

- 零阶项为零（相同分布的 KL 散度为零）
- 一阶项为零：将概率归一化条件 $\sum_x p_{\boldsymbol{\theta}}(x)=1$ 对 θ^j 求导给出

$$\partial_j\sum_x p_{\boldsymbol{\theta}}(x)\overset{\text{求导积分换序}}{=}\sum_x\partial_j p_{\boldsymbol{\theta}}(x)=\sum_x p_{\boldsymbol{\theta}}(x)\,\partial_j\ln p_{\boldsymbol{\theta}}(x)=0. \tag{7}$$

  即一阶项也为零。这里隐含了一个条件：对于 $p_{\boldsymbol{\theta}}(x)$，求导和积分可以换序，也即**正则化条件**。

- 二阶项：对式 (7) 再关于 θ^i 求导，

$$\begin{aligned}
0&=\partial_i\sum_x p_{\boldsymbol{\theta}}(x)\,\partial_j\ln p_{\boldsymbol{\theta}}(x)\\
&=\sum_x\partial_i\left[p_{\boldsymbol{\theta}}(x)\,\partial_j\ln p_{\boldsymbol{\theta}}(x)\right]\\
&=\sum_x\left[p_{\boldsymbol{\theta}}(x)\,\partial_i\ln p_{\boldsymbol{\theta}}(x)\cdot\partial_j\ln p_{\boldsymbol{\theta}}(x)+p_{\boldsymbol{\theta}}(x)\,\partial_i\partial_j\ln p_{\boldsymbol{\theta}}(x)\right].
\end{aligned}$$

  这里隐含的条件是对于 $\partial_j\ln p_{\boldsymbol{\theta}}(x)$，求导和积分可以换序。移项得

$$\sum_x p_{\boldsymbol{\theta}}(x)\,\partial_i\ln p_{\boldsymbol{\theta}}(x)\cdot\partial_j\ln p_{\boldsymbol{\theta}}(x)=-\sum_x p_{\boldsymbol{\theta}}(x)\,\partial_i\partial_j\ln p_{\boldsymbol{\theta}}(x). \tag{9}$$

  将式子左边定义为 **Fisher 信息矩阵**

$$I_{ij}(\boldsymbol{\theta}):=\sum_x p_{\boldsymbol{\theta}}(x)\,\partial_i\ln p_{\boldsymbol{\theta}}(x)\cdot\partial_j\ln p_{\boldsymbol{\theta}}(x). \tag{10}$$

观察到式子右边是 $\ln p_{\boldsymbol{\theta}}(x)$ 对应的 Hessian 矩阵的负期望值，即

$$\text{Fisher 信息矩阵} = -\mathbb{E}_{p_{\boldsymbol{\theta}}}[\text{Hessian of } \ln p_{\boldsymbol{\theta}}(x)] \tag{2}$$

代入前述 KL 散度的泰勒展开，得到

$$D_{\mathrm{KL}}(p_{\boldsymbol{\theta}} \| p_{\boldsymbol{\theta}+\delta\boldsymbol{\theta}}) \approx \frac{1}{2}\delta\theta^i \delta\theta^j I_{ij}(\boldsymbol{\theta}) = \frac{1}{2}\delta\boldsymbol{\theta}^\top I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta}. \tag{11}$$

这表明在局部我们可以用 I_{ij} 作为距离来近似 KL 散度。这是 KL 散度诱导的一个黎曼度规，称为 **Fisher-Rao 度规**，从而我们可以将分布们看作一个黎曼流形，并谈论其上的距离。

表 1: KL 散度泰勒展开各阶项

阶数	值	推导	条件
0	0	$D_{\mathrm{KL}}(p\|p) = 0$	—
1	0	$\partial_i\big(\sum_x p_{\boldsymbol{\theta}}(x) - 1\big) = 0$	积分、求导可换序
2	$\frac{1}{2}\delta\boldsymbol{\theta}^\top I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta}$	$\partial_i\partial_j\big(\sum_x p_{\boldsymbol{\theta}}(x) - 1\big) = 0$	积分、求导可换序

1.2 Score 向量

为了表述方便，我们定义 **score 向量**（对数似然梯度）

$$s_i(x) := \partial_i \ln p_{\boldsymbol{\theta}}(x),$$

并引入期望记号 $\mathbb{E}_{p_{\boldsymbol{\theta}}}[f(x)] := \sum_x p_{\boldsymbol{\theta}}(x)\, f(x)$。则前述各式可简写为：

- 式 (7)：$\mathbb{E}[s_i] = 0$（score 均值为零）；
- 式 (10)：$I_{ij} = \mathbb{E}[s_i\, s_j]$（Fisher 信息 = score 外积的期望）。

由于 $\mathbb{E}[s_i] = 0$，I_{ij} 同时也是 score 的协方差矩阵：$I_{ij} = \mathrm{Cov}(s_i, s_j)$。

实践提醒：上述期望取自模型分布 $p_{\boldsymbol{\theta}}$。在机器学习中，损失函数的 Hessian $H_{ij} = \partial_i\partial_j\mathcal{L}$ 一般不等于 Fisher 信息矩阵。二者相等当且仅当：(1) 损失为负对数似然 $\mathcal{L} = -\ln p_{\boldsymbol{\theta}}(x)$；(2) 期望取自模型分布而非数据分布。

1.3 流形上的最速下降

考虑优化损失函数 $\mathcal{L}(\boldsymbol{\theta})$，其梯度为 $g = \nabla_{\boldsymbol{\theta}}\mathcal{L}$。普通梯度下降以欧氏范数 $\|\delta\boldsymbol{\theta}\|_2$ 约束步长，但这不具备重参数化不变性。若改用 Fisher-Rao 度规 $\|\delta\boldsymbol{\theta}\|_I^2 = \delta\boldsymbol{\theta}^\top I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta}$ 约束步长，则最速下降问题为

$$\max_{\delta\boldsymbol{\theta}} \; -g^\top \delta\boldsymbol{\theta} \quad \text{s.t.} \quad \delta\boldsymbol{\theta}^\top I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = \epsilon^2.$$

用 Lagrange 乘子法，得

$$I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = -\eta g, \quad \Rightarrow \quad \delta\boldsymbol{\theta} = -\eta I(\boldsymbol{\theta})^{-1} g. \tag{12}$$

这就是 Amari 提出的自然梯度（Natural Gradient Descent, NGD）更新公式。

2 从经典概率分布到量子态

2.1 经典模型 vs. 量子模型：输出与冗余

经典概率模型直接给出实非负、归一化的概率分布 $p_\theta(x)$。量子模型（例如参数化变分 Ansatz）返回态向量 $|\Psi(\theta)\rangle$，概率通过 Born 规则

$$P(x) = \frac{|\langle x|\Psi(\theta)\rangle|^2}{\langle\Psi(\theta)|\Psi(\theta)\rangle}.$$

乘以任意非零复数 c 不会改变上式，因此 $|\Psi\rangle \sim c|\Psi\rangle$（全局相位与整体幅值）构成 **规范冗余**。

2.2 Fisher 信息为何不能直接照搬到复振幅

若把复振幅当作"输出"并生搬 Fisher 信息，会遇到：

1. 振幅不可直接观测，无法在概率空间中定义期望；
2. 全局相位方向不影响概率，却令信息矩阵沿该方向奇异；
3. 整体归一化同理，也会被误计入。

因此须先除去规范自由度，再谈度量。

2.3 正确路线：Fubini–Study 距离 ⇒ 量子几何张量

纯态的物理空间是射影 Hilbert 空间 $\mathbb{CP}^{n-1}$。其天然距离是 *Fubini–Study (FS) 距离*。对 FS 距离在参数空间做二阶展开产生 **量子几何张量** (QGT)：实部给出 FS 黎曼度量，虚部给出 Berry 曲率——与经典"KL ⇒ Fisher"完全对应。

3 Fubini-Study 距离与量子几何张量（QGT）：规范不变定义与推导

3.1 Fubini-Study 距离：射影 Hilbert 空间上的天然距离

对于两个（不必归一化的）态 $|\Psi\rangle$ 与 $|\Phi\rangle$，Fubini-Study 距离定义为

$$d_{\mathrm{FS}}(|\Psi\rangle, |\Phi\rangle) = \arccos \frac{|\langle\Psi|\Phi\rangle|}{\sqrt{\langle\Psi|\Psi\rangle\langle\Phi|\Phi\rangle}}. \tag{13}$$

显然 d_{FS} 对 $|\Psi\rangle \to c|\Psi\rangle$ 和 $|\Phi\rangle \to c'|\Phi\rangle$ 不变，因此是定义在射影空间上的距离。

3.2 从 FS 距离到 QGT：二阶展开的详细计算

设 $|\Psi(\theta)\rangle$ 是参数化的态，考虑 $|\Psi(\theta)\rangle$ 与 $|\Psi(\theta+\delta\theta)\rangle$ 的 FS 距离平方。记

$$F(\delta\theta) = \frac{|\langle\Psi(\theta)|\Psi(\theta+\delta\theta)\rangle|^2}{\langle\Psi(\theta)|\Psi(\theta)\rangle\langle\Psi(\theta+\delta\theta)|\Psi(\theta+\delta\theta)\rangle}.$$

在 $\delta\theta = 0$ 处 $F(0) = 1$，展开 F 到二阶：

$$F(\delta\theta) \approx 1 - \delta\theta^\mu \delta\theta^\nu \Re(Q_{\mu\nu}) + O(\|\delta\theta\|^3),$$

其中 $Q_{\mu\nu}$ 即为量子几何张量。由 $d_{\mathrm{FS}}^2 \approx 2(1-\sqrt{F})$ 可得

$$\mathrm{d}s_{\mathrm{FS}}^2 = g_{\mu\nu}\,\mathrm{d}\theta^\mu \mathrm{d}\theta^\nu, \quad g_{\mu\nu} = \Re(Q_{\mu\nu}). \tag{14}$$

3.3 QGT 的规范不变定义与分解：度量 + Berry 曲率

对于依赖参数 θ 的（不必归一化的）纯态 $|\Psi(\theta)\rangle$，QGT 定义为

$$Q_{\mu\nu} = \frac{\langle\partial_\mu\Psi|(I-\Pi_\Psi)|\partial_\nu\Psi\rangle}{\langle\Psi|\Psi\rangle}, \quad \Pi_\Psi = \frac{|\Psi\rangle\langle\Psi|}{\langle\Psi|\Psi\rangle}. \tag{15}$$

其中 $|\partial_\mu\Psi\rangle \equiv \partial_{\theta^\mu}|\Psi(\theta)\rangle$。投影算符 $I-\Pi_\Psi$ 扣除了导数态在 $|\Psi\rangle$ 方向的分量，保证规范不变性。

展开式 (15) 可得

$$Q_{\mu\nu} = \frac{\langle\partial_\mu\Psi|\partial_\nu\Psi\rangle}{\langle\Psi|\Psi\rangle} - \frac{\langle\partial_\mu\Psi|\Psi\rangle\langle\Psi|\partial_\nu\Psi\rangle}{\langle\Psi|\Psi\rangle^2}. \tag{16}$$

$Q_{\mu\nu}$ 是复值张量，可分解为

$$Q_{\mu\nu} = g_{\mu\nu} + \frac{i}{2}F_{\mu\nu}, \quad g_{\mu\nu} = \Re(Q_{\mu\nu}), \quad F_{\mu\nu} = 2\Im(Q_{\mu\nu}). \tag{17}$$

其中对称的实部 $g_{\mu\nu}$ 是 Fubini-Study 度量在参数空间的表示，反对称的虚部 $F_{\mu\nu}$ 是 Berry 曲率。

3.4 测量基表示与 Monte Carlo 估计：QGT 作为" 对数导数的协方差"

在计算基 $\{|x\rangle\}$ 下，定义 $\Psi(x) = \langle x|\Psi\rangle$ 及对数导数

$$O_\mu(x) = \frac{\partial_\mu\Psi(x)}{\Psi(x)} = \partial_\mu \ln\Psi(x). \tag{18}$$

则 QGT 可写为

$$Q_{\mu\nu} = \mathbb{E}_P[O_\mu^* O_\nu] - \mathbb{E}_P[O_\mu^*]\mathbb{E}_P[O_\nu], \tag{19}$$

其中期望取自 Born 概率 $P(x) = |\Psi(x)|^2/\langle\Psi|\Psi\rangle$。这一形式与经典 Fisher 信息的"score 协方差" 形式完全类比，但注意这里的 O_μ 是复值。

4 随机重构 SR 与最速下降：FS 范数约束下的严格推导，并与经典自然梯度对照

4.1 SR 的基本线性方程

随机重构（Stochastic Reconfiguration, SR）是变分蒙特卡洛中常用的参数更新方法，其核心方程为

$$S(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = -\eta g, \tag{20}$$

其中 $S(\boldsymbol{\theta})$ 是 QGT 的实部（或其对称化/正则化版本），$g = \nabla_{\boldsymbol{\theta}} E$ 是能量梯度。

4.2 证明：FS 范数约束的最速下降推出 SR

考虑优化能量 $E(\boldsymbol{\theta}) = \langle\Psi(\boldsymbol{\theta})|\hat{H}|\Psi(\boldsymbol{\theta})\rangle/\langle\Psi|\Psi\rangle$，其梯度为 $g = \nabla_{\boldsymbol{\theta}} E$。以 FS 度量约束步长范数：

$$\max_{\delta\boldsymbol{\theta}} \; -g^\top\delta\boldsymbol{\theta} \quad \text{s.t.} \quad \delta\boldsymbol{\theta}^\top S(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = \epsilon^2.$$

用 Lagrange 乘子法，令

$$\mathcal{L}(\delta\boldsymbol{\theta}, \lambda) = -g^\top\delta\boldsymbol{\theta} + \frac{\lambda}{2}(\delta\boldsymbol{\theta}^\top S\,\delta\boldsymbol{\theta} - \epsilon^2).$$

对 $\delta\boldsymbol{\theta}$ 求导并令其为零：

$$-g + \lambda S\,\delta\boldsymbol{\theta} = 0 \quad \Rightarrow \quad S\,\delta\boldsymbol{\theta} = -\frac{1}{\lambda}g.$$

将 $1/\lambda$ 吸收到学习率 η 中，即得 SR 方程 (20)。

4.3 与经典自然梯度的严格类比

比较经典自然梯度 $I(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = -\eta g$ 与量子 SR $S(\boldsymbol{\theta})\,\delta\boldsymbol{\theta} = -\eta g$，可见两者具有完全相同的形式，只是度量张量不同：

- 经典：$M = I(\boldsymbol{\theta})$（Fisher 信息矩阵）← KL 散度二阶展开 ← 概率分布流形；
- 量子：$M = S(\boldsymbol{\theta})$（QGT 实部）← FS 距离二阶展开 ← 射影 Hilbert 空间。

4.4 数值实现要点：奇异性、正则化与对称化

由于参数冗余或规范自由度，S 往往病态或半正定而不可逆。实践中常见处理包括：

- Tikhonov 正则化：$S \to S + \lambda I$；
- 截断/伪逆：对小特征值截断；
- 对称化：用 $\frac{1}{2}(S + S^{\top})$ 或取实部 $\Re S$ 以保证对称半正定。

这些不改变几何主干：SR 的本质是" 用内禀度量修正梯度方向"。

表 2: 经典与量子信息几何的关键对应关系

概念	经典概率模型	量子纯态模型
" 状态" 对象	概率分布 $p_\theta(x)$	态向量/射线 $\lvert\Psi(\theta)\rangle \sim c\lvert\Psi(\theta)\rangle$
输出的可观测量	概率本身可观测（抽样）	振幅不可直接观测；概率由 $P(x) = \lvert\Psi(x)\rvert^2/\langle\Psi\lvert\Psi\rangle$ 给出
冗余/规范	主要是重参数化（坐标变换）	既有重参数化，也有全局相位/整体尺度等规范冗余
" 距离/散度"	KL 散度 $D_{\mathrm{KL}}(p_\theta \Vert p_{\theta+\delta})$	FS 距离 $\mathcal{D}_{\mathrm{FS}}(\psi,\psi') = \arccos\lvert\langle\psi\lvert\psi'\rangle\rvert$
局部二阶展开	$D_{\mathrm{KL}} \approx \frac{1}{2}\delta\theta^{\top} I \delta\theta$	$\mathrm{d}s^2_{\mathrm{FS}} = \delta\theta^{\top} g \delta\theta$（差常数因子）
黎曼度规	Fisher 信息矩阵 $I_{ij} = \mathbb{E}[s_i s_j]$	QGT $Q_{\mu\nu} = \langle\partial_\mu\psi\lvert(I - \lvert\psi\rangle\langle\psi\rvert)\partial_\nu\psi\rangle$
" 协方差" 表述	score 协方差（对 p 取期望）	log-derivative 协方差（对 $\lvert\Psi\rvert^2$ 取期望）+ 规范投影
额外的反对称结构	无（常见 FIM 为实对称）	Berry 曲率 $F_{\mu\nu} = 2\Im Q_{\mu\nu}$
最速下降更新	固定 Fisher 范数 ⇒ 自然梯度 $I\delta\theta = -\eta\nabla J$	固定 FS 范数 ⇒ SR/QNG $S\delta\theta = -\eta\nabla E$

4.5 表 2：度量、优化方法与流形视角（含 Adam/RMSProp 的对角近似）

结语

本文把经典的 Fisher-Rao 信息几何与量子态的 FS/QGT 几何放在同一框架下：先选对" 距离/散度"，再用二阶展开得到度量，并用该度量定义最速下降。经典的 KL ⇒ Fisher ⇒ 自然梯度；量子的 FS ⇒ QGT ⇒ SR/QNG。从算法角度看，许多优化方法都可理解为在更新方程 $M(\theta)\delta\theta = -\eta g$ 中选择不同的 M：从欧氏单位阵到 Fisher/QGT，再到各种对角/结构化近似。

表 3: 优化算法、度量/预条件与流形解释的统一视角

方法	**隐含度量/预条件** $M(\theta)$	**流形/解释**
SGD / GD	$M = I$（欧氏）	把参数空间当作欧氏空间; 步长与坐标有关，不具备重参数化不变性。
动量法	$M \approx I+$ 动量累积	仍是欧氏几何，但用一阶滤波改善病态方向的振荡。
自然梯度 (NGD)	$M = I(\theta)$ (FIM)	统计流形 (Fisher-Rao 度量); 等价于固定 KL 二阶近似的最速下降。
K-FAC / GGN 类方法	$M \approx$ (FIM 或 GGN 的结构化近似)	近似捕捉参数耦合的曲率/信息；在深度网络中常用块对角/克罗内克分解近似。
RMSProp / AdaGrad	$M = \mathrm{diag}(v)$（梯度二阶矩的对角）	经验上可看作对角预条件：当损失为 NLL 且在模型分布上取期望时，$\mathbb{E}[g_i^2]$ 与 FIM 对角元同阶，因此近似" 对角 Fisher"。
Adam	$M = \mathrm{diag}(\hat{v})$（带偏差校正的一阶/二阶矩对角）	可视为 RMSProp + 动量; 依然是对角近似（忽略参数相关性），但数值鲁棒且易用。
牛顿法	$M = \nabla^2 J(\theta)$ (Hessian)	在欧氏参数空间用二阶曲率；非凸时 Hessian 可不定，需阻尼/近似。
SR / QNG	$M = S \approx g_{\mathrm{FS}}$（QGT 实部/对称化）	射影 Hilbert 空间 (FS 度量)；等价于固定 FS 范数的最速下降 (TDVP 离散化)。
镜像下降	M 由 Bregman 散度诱导	一般化" 先选散度再做最速下降" 的框架；KL 与 FS 都是其中的重要实例。

参考文献

[1] S.-i. Amari, *Natural Gradient Works Efficiently in Learning*, Neural Computation **10**(2), 251–276 (1998).

[2] S. Sorella, *Green Function Monte Carlo with Stochastic Reconfiguration*, Phys. Rev. Lett. **80**, 4558 (1998).

[3] S. Sorella, *Wave function optimization in the variational Monte Carlo method*, Phys. Rev. B **71**, 241103(R) (2005).

[4] J. Provost and G. Vallée, *Riemannian structure on manifolds of quantum states*, Commun. Math. Phys. **76**, 289–301 (1980).

[5] J. Izaac, C. Wang, and Z. Wang, *Quantum Natural Gradient*, arXiv:1811.08451 (2019).