

一种无导数优化方法及其应用

南京大学本科毕业论文答辩

指导老师：宋凤麒教授

答辩人：张楚珩 南京大学物理学院

答辩日期：二零一六年六月二日

概要

- 优化方法简介
- 一种无导数优化方法——Racos
- Racos在光学器件设计上的应用
- Racos在神经网络训练上的应用

优化方法简介

- 优化问题具有形式

$$\mathit{arg} \min_x f(\boldsymbol{x})$$

subject to $\boldsymbol{x} \in \Omega$

- 优化问题具有广泛的应用前景
- 优化方法用于解决优化问题

优化方法简介

- 根据在优化过程中是否使用导数可以分为**基于导数的优化方法**和**无导数优化方法**

	基于导数的优化方法	无导数优化方法
优点	效率高	应用于导数难获取的情形 避免导数误导，跳出局部极小
缺点	有时候导数信息不可获得 或者获取成本很高 容易陷入局部极小	效率较基于导数的方法低

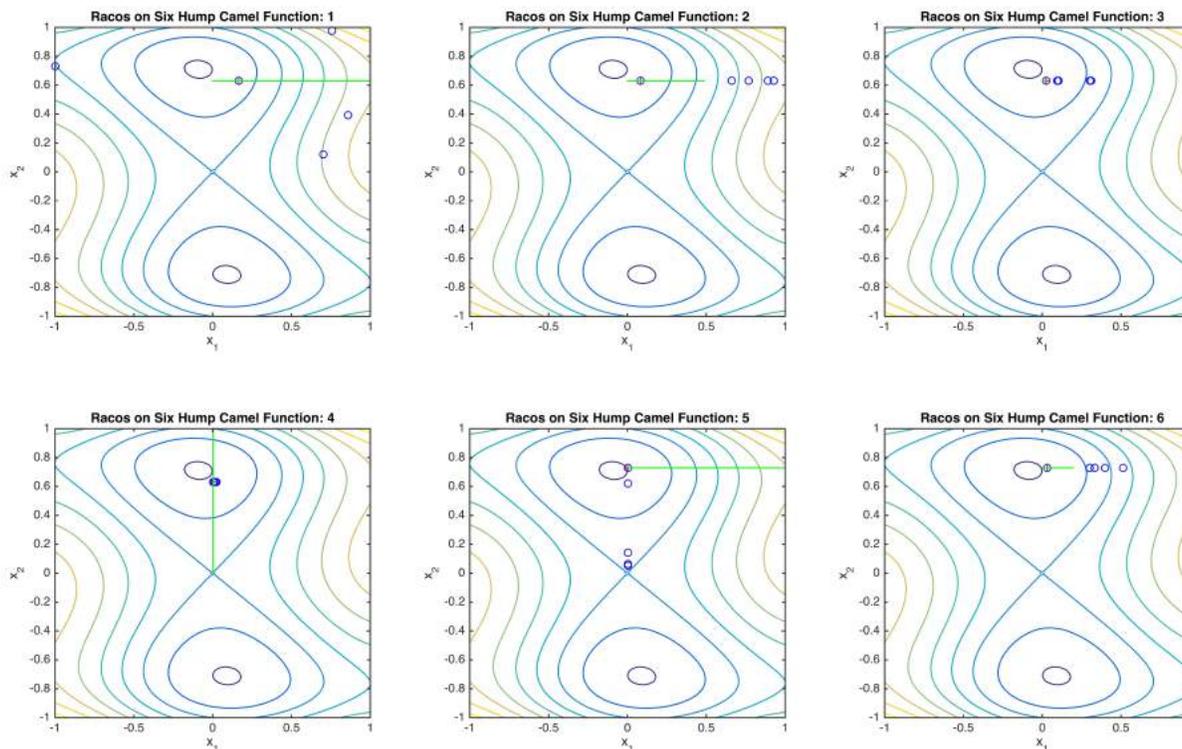
光学器件设计

神经网络训练

一种无导数优化方法——Racos

主要思想:

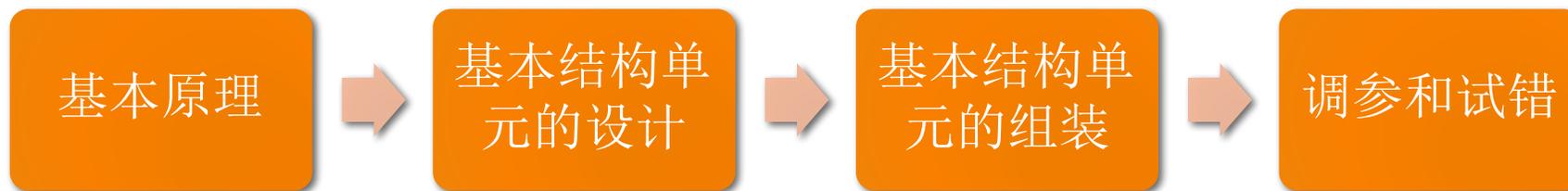
- 随机采点
- 区分好坏
- 估计函数
- 针对采样



Racos算法在六驼峰函数上的优化。蓝色圆点为每一代的采样点，红色十字为被标记出的“好”采样点，绿色方框为分类算法给出来的假设。每一轮迭代中，Racos对种群进行采样并标记，并根据标记结果给出假设，使用假设来确定下一轮的采样点。

Racos在光学器件设计上的应用：原理

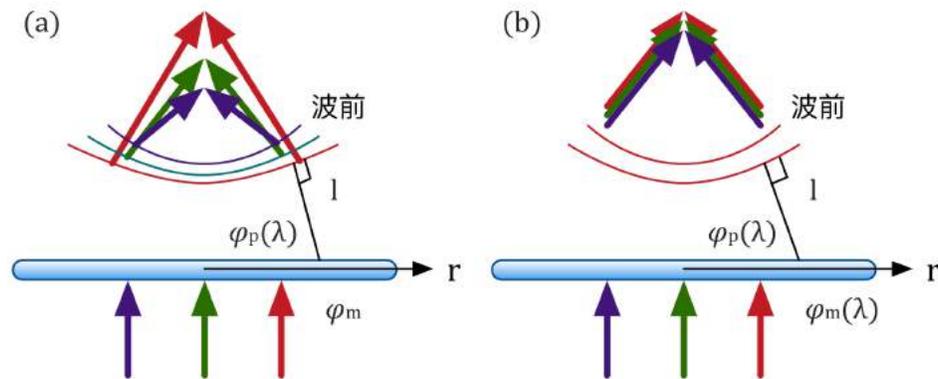
正向设计：



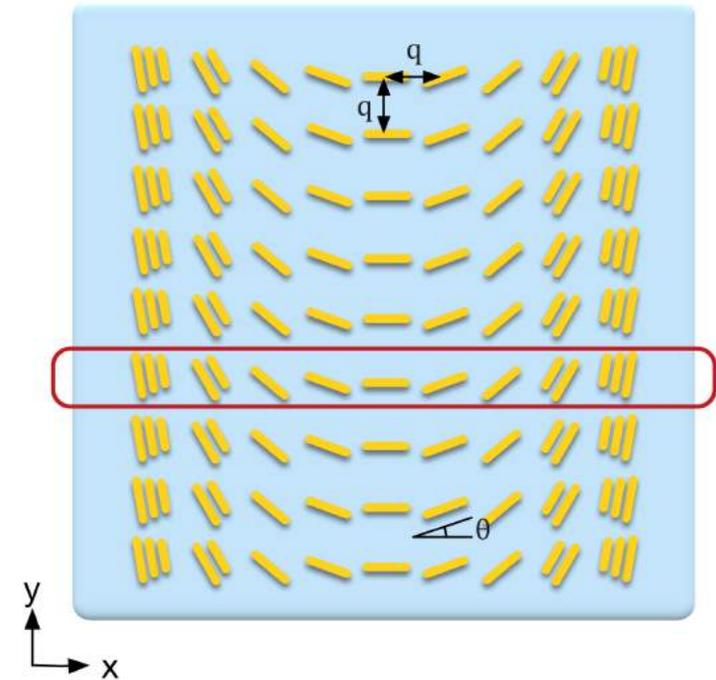
逆向设计：



Racos在光学器件设计上的应用： 消色差超透镜

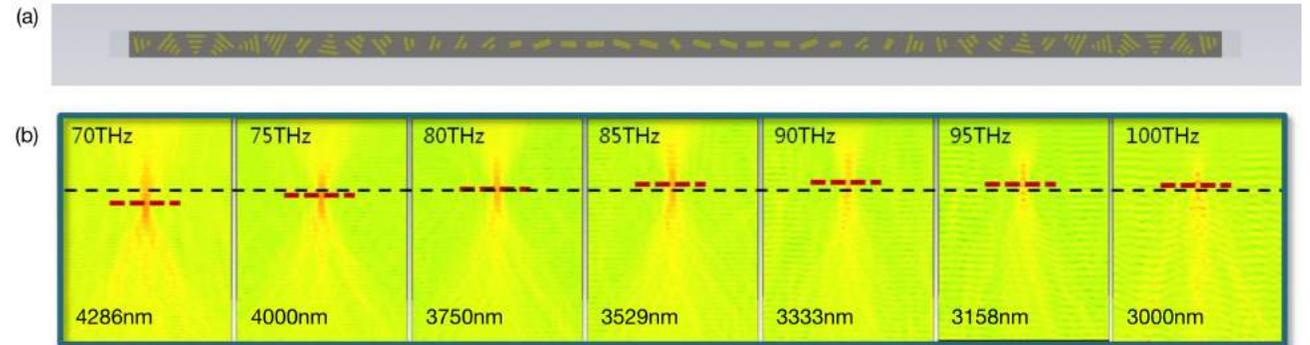
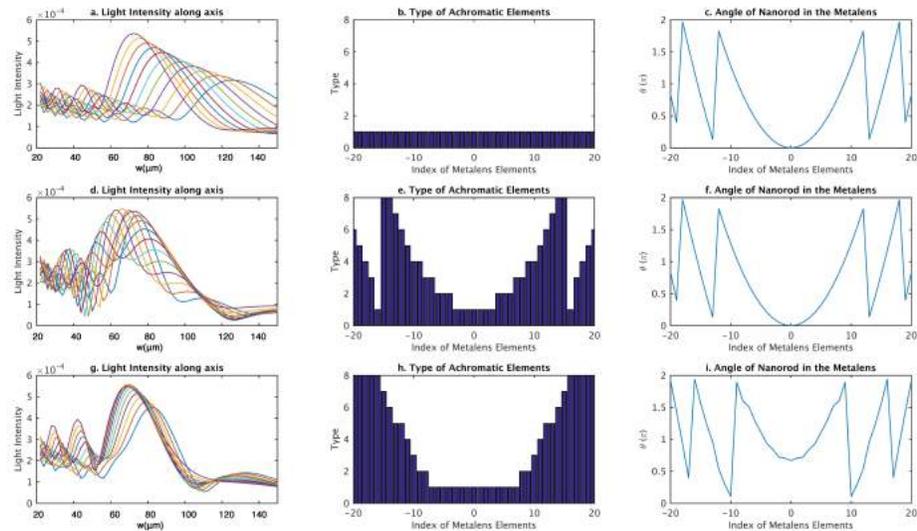


消色差超透镜的基本原理。 (a) 如果超透镜对于任何波长的光都附加相同的相位，不同补偿的光将无法聚焦到同一点，产生色散。(b) 按照一定的规律设计超透镜各个位置对于不同波长的相位补偿可以消除色散。



超透镜结构。 超透镜由介质衬底和上面的金属小棒阵列构成。每个金属小棒单元拥有不同的转角，不同数目和大小的小棒，以及不等的小棒间间距。

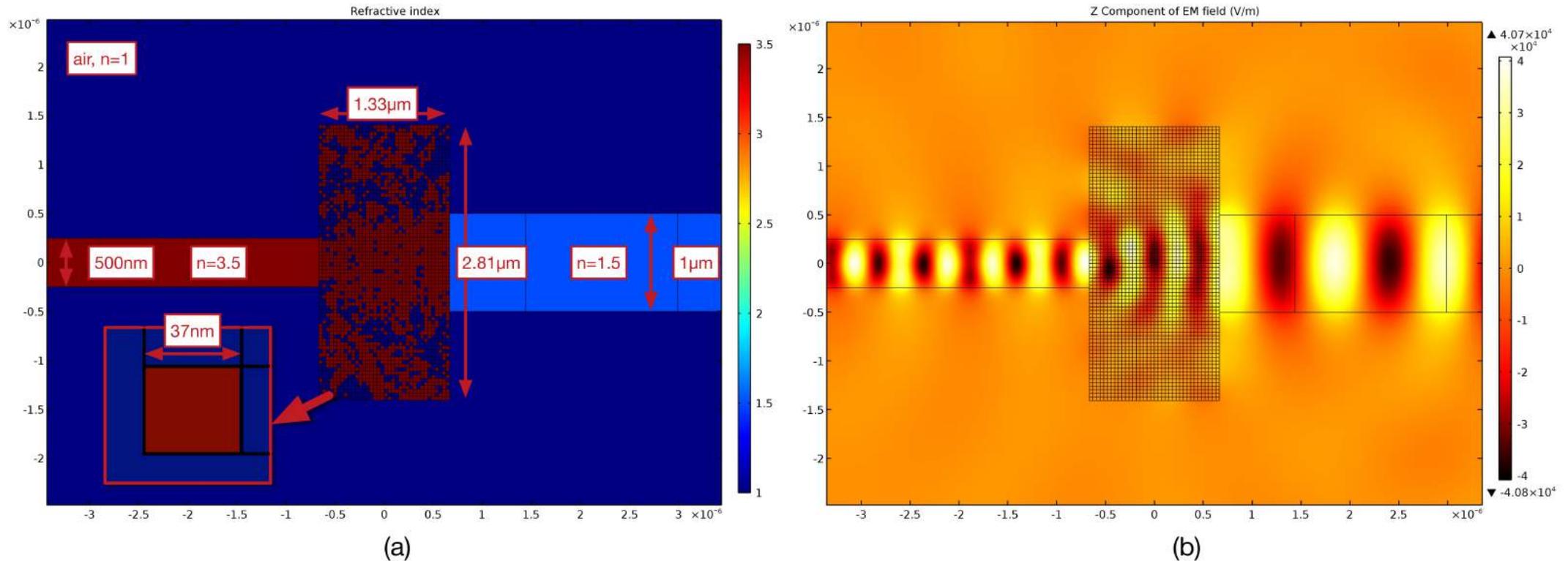
Racos在光学器件设计上的应用： 消色差超透镜的设计结果



超透镜消色差优化结果。三行分别表示1) 无消色差设计下超透镜的仿真结果；2) 正向设计下的消色差超透镜仿真结果；3) 逆向设计下的消色差超透镜仿真结果。三列分别表示1) 沿中轴线的场强分布；2) 消色差部分的设计；3) 转角部分的设计。

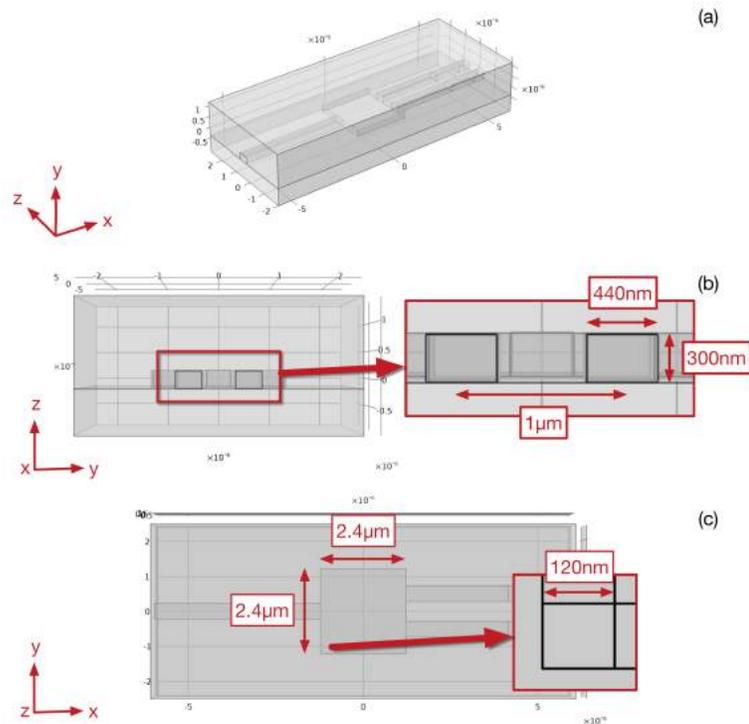
消色差超透镜的模拟结果。(a) 超透镜柱镜设计方案。(b) 3000nm~4286nm波段内聚焦涨落在 $\pm 5.9\%$ 以内，红色虚线表示各波长聚焦位置，黑色虚线表示平均聚焦位置。

Racos在光学器件设计上的应用： 波导换模器的设计结果

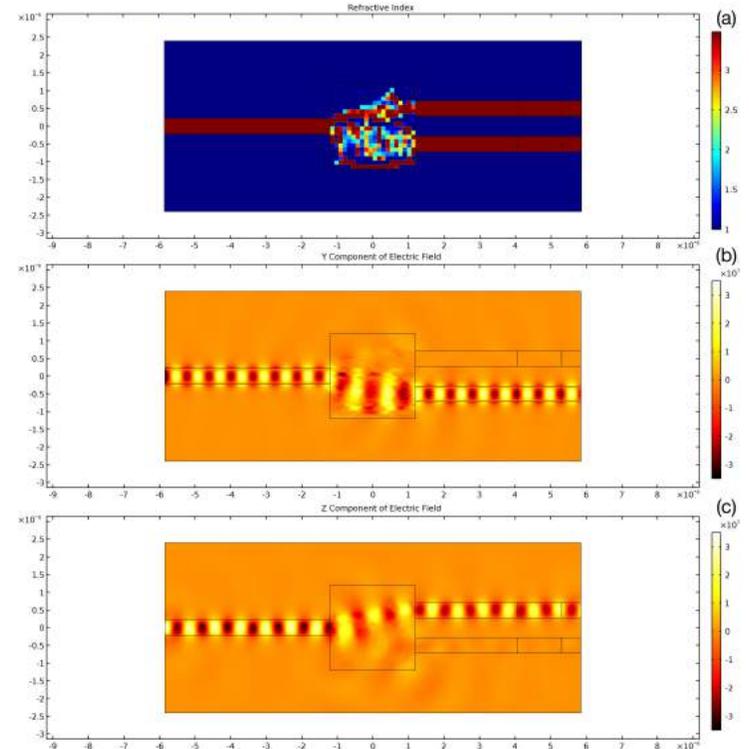


换模器设计方案与模拟结果。(a)换模器的设计方案，其设计区域为长 $1.33\mu\text{m}$ ，宽 $2.81\mu\text{m}$ 的区域，每个像素尺寸 $37\text{nm} \times 37\text{nm}$ ，工作波长为 $\lambda = 1550\text{nm}$ 。(b)换模器的模拟结果。

Racos在光学器件设计上的应用： 硅基偏振分束器的设计结果



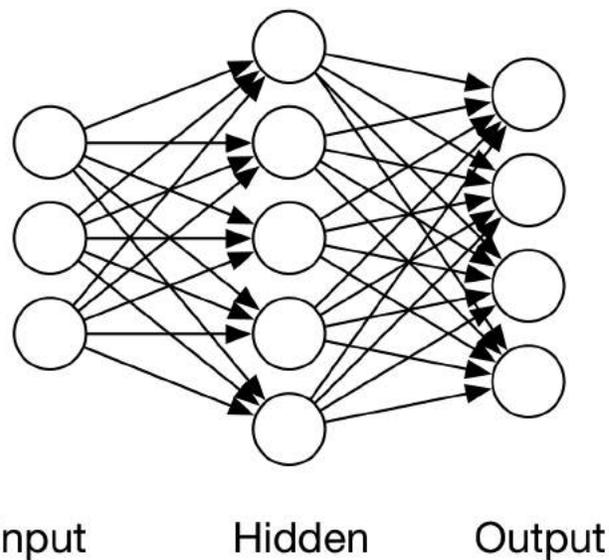
偏振分束器几何结构和尺寸。波导横截面宽 440nm ，高 300nm ，出射波导管间距 $1\mu\text{m}$ ，偏振分束器大小 $2.4\mu\text{m}\times 2.4\mu\text{m}$ ，每个像素大小 $120\text{nm}\times 120\text{nm}$ ，偏振分束器工作波长为 1550nm 。



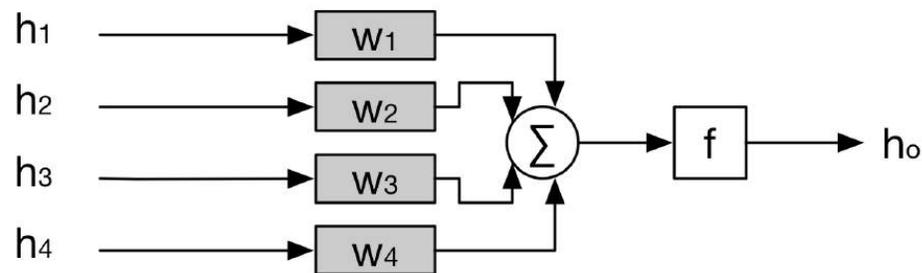
偏振分束器设计方案与模拟结果。(a)偏振分束器的设计方案；(b)入射波导模式为TM模式时的电场分布；(c)入射波导模式为TE模式时的电场分布。

Racos在神经网络训练上的应用：原理

BP神经网络简介



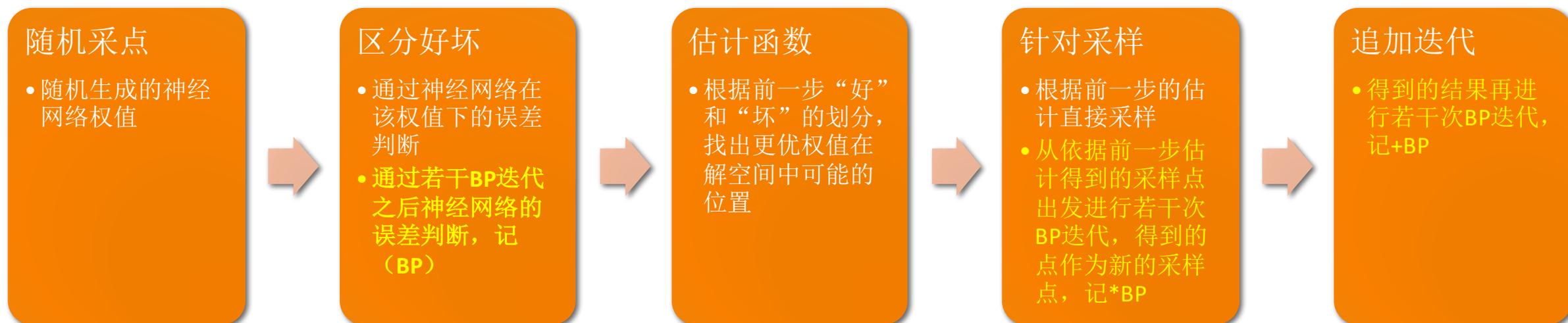
BP神经网络的结构。它包括输入层、隐含层和输出层；每一层都含有数目不同的神经元，每一个神经元都与相邻层的所有神经元相连。



BP神经网络中神经元的结构。前一层的信号分别于相应权值相乘后累加，再通过一个非线性映射得到该神经元的输出。

Racos在神经网络训练上的应用：策略

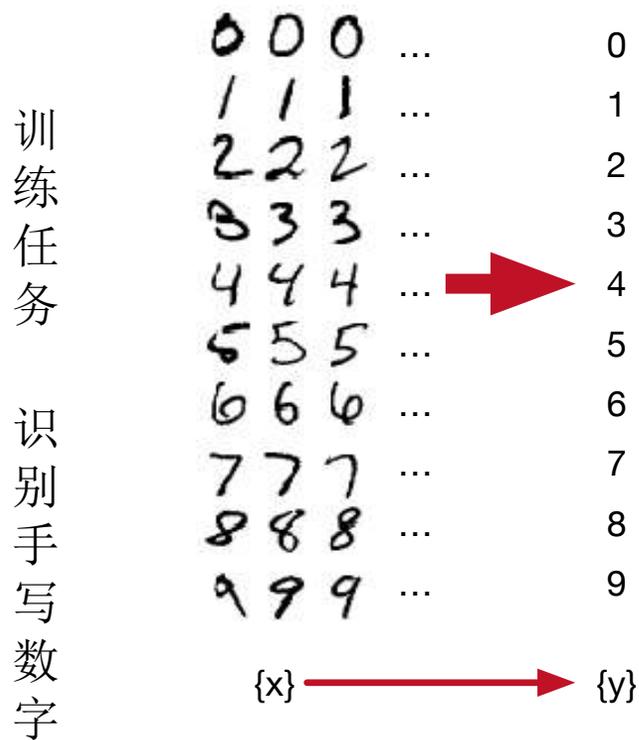
Racos与BP算法的混合策略



混合策略：Racos(BP)、Racos*BP、Racos+BP、Racos(BP)+BP、Racos*BP+BP

对照算法：BP、Racos

Racos在神经网络训练上的应用：结果



MNIST 数据集。它的样本输入为大量手写数字的图片, 目标输出为每张图片所对应的数字。

算法	误差率
<i>Racos</i>	22.044%
<i>Racos*BP</i>	1.9094%
<i>Racos(BP)</i>	3.9317%
<i>Racos+BP</i> ($\eta = 0.1$)	1.5967%
<i>Racos*BP+BP</i> ($\eta = 0.1$)	1.4634%
<i>Racos(BP)+BP</i> ($\eta = 0.1$)	1.5790%
<i>BP</i>	2.2825%

各个算法在固定30000单位计算成本下的误差率。*Racos*BP+BP*在此数据集上有最好的训练效果, *Racos*BP+BP*、*Racos(BP)+BP*、*Racos*BP*和*Racos+BP*都比仅仅基于梯度的*BP*方法有更好的效果。

重要参考文献

- [1] Yu, Yang, Hong Qian, and Yi-Qi Hu. "Derivative-free optimization via classification." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- [2] Chong, Edwin KP, and Stanislaw H. Zak. *An introduction to optimization*. Vol. 76. John Wiley & Sons, 2013.
- [3] Anzai, Yuichiro. *Pattern Recognition & Machine Learning*. Elsevier, 2012.
- [4] Zhi-Hua Zhou. *Machine Learning*. Tsinghua University Press, 2016.
- [5] Shen, Bing, et al. "An integrated-nanophotonics polarization beamsplitter with $2.4 \times 2.4 \mu\text{m}^2$ footprint." *Nature Photonics* 9.6 (2015): 378-382.
- [6] Chen, Xianzhong, et al. "Dual-polarity plasmonic metalens for visible light." *Nature communications* 3 (2012): 1198.
- [7] Aieta, Francesco, et al. "Multiwavelength achromatic metasurfaces by dispersive phase compensation." *Science* 347.6228 (2015): 1342-1345.
- [8] Lu, Jesse, and Jelena Vučković. "Nanophotonic computational design." *Optics express* 21.11 (2013): 13351-13367.

致谢

感谢俞扬副教授、王湫明研究员和宋凤麒教授在毕业设计上给予我的支持和帮助！

感谢在场评各位审老师的耐心聆听！

提问： 消色差透镜设计相关公式

目标函数

$$f(\mathbf{x}) = - \int_{\lambda_1}^{\lambda_2} \int_{f-\sigma}^{f+\sigma} |E(\mathbf{x}, w, \lambda)|^2 dw d\lambda + \epsilon \int_{\lambda_1}^{\lambda_2} \left| \int_{f-\sigma}^{f+\sigma} |E(\mathbf{x}, w, \lambda)|^2 dw - \bar{E}^2 \right|^2 d\lambda$$

场强

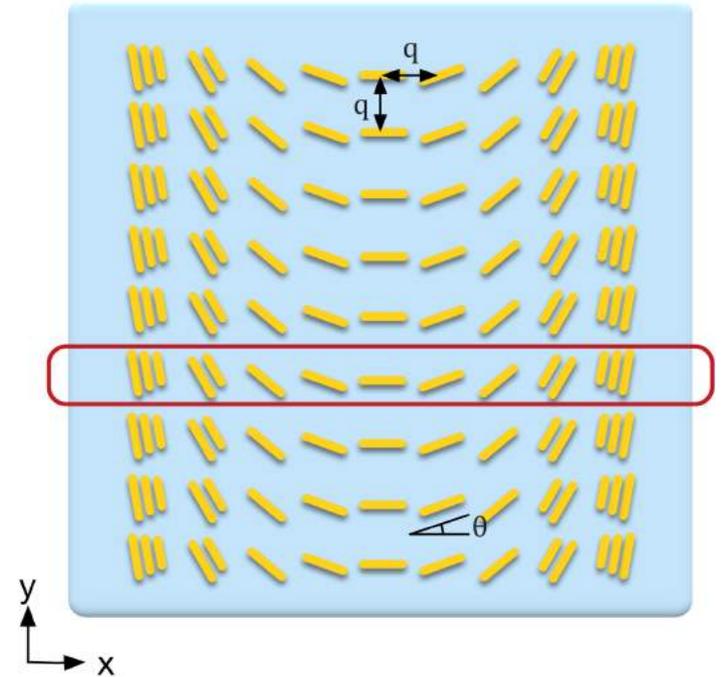
$$E(\mathbf{x}, w, \lambda) = c \int_{-r_0}^{r_0} \exp\left(j \frac{2\pi}{\lambda} \sqrt{w^2 + r^2} + j\varphi_m(r)\right) dr$$

提问： 消色差透镜基础附加相位的计算

$$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \alpha_e \begin{pmatrix} \cos^2 \varphi & \sin \varphi \cos \varphi \\ \sin \varphi \cos \varphi & \sin^2 \varphi \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix}$$

$$\begin{aligned} \mathbf{P}_{L(R)} &= p_x \mathbf{e}_x + ip_y \mathbf{e}_y \\ &= \frac{1}{2} \alpha_e E(\mathbf{e}_x \pm i \mathbf{e}_y) + \frac{1}{2} \alpha_e e^{\pm i 2 \varphi} E(\mathbf{e}_x \mp i \mathbf{e}_y) \\ &= \frac{1}{\sqrt{2}} \alpha (\mathbf{e}_{L(R)} \pm e^{\pm i 2 \varphi} \mathbf{e}_{R(L)}) \end{aligned}$$

$$\mathbf{E}_{L(R)} = E(\mathbf{e}_x \pm i \mathbf{e}_y) / \sqrt{2} \quad \mathbf{e}_{L(R)} = (\mathbf{e}_x \pm i \mathbf{e}_y) / \sqrt{2}$$



提问： 微纳光子学设计相关公式

$$\nabla \times \mu_0^{-1} \nabla \times \mathbf{E} - \omega^2 \epsilon \mathbf{E} = -i\omega \mathbf{J}$$

$$\epsilon \rightarrow \mathbf{x}, (\nabla \times \mu_0^{-1} \nabla \times) + (-\omega^2 \epsilon) \rightarrow A(\mathbf{x}), -i\omega \mathbf{J} \rightarrow \mathbf{b}$$

$$A(\mathbf{x}) \mathbf{E} = \mathbf{b}$$

$$\arg \min_{\mathbf{x}} \int_{\text{probe region}} \|\mathbf{E} - \hat{\mathbf{E}}\|^2 d\sigma$$

$$\text{subject to } A(\mathbf{x}) \mathbf{E} = \mathbf{b}, \mathbf{x}_{\min} \leq \mathbf{x} \leq \mathbf{x}_{\max}$$

波导换模器

$$\arg \max_{\mathbf{x}} \int_{\text{probe region}} \|\mathbf{E}\|^2 d\sigma$$

$$\text{subject to } A(\mathbf{x}) \mathbf{E} = \mathbf{b}, \mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2\}$$

偏振分束器

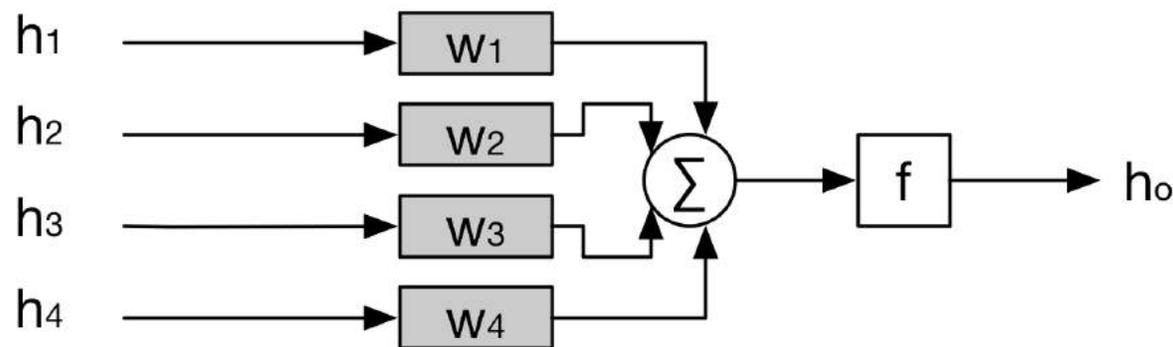
$$\arg \max_{\mathbf{x}} \left(\int_{\text{probe1}} \|\mathbf{E}_{TE}\|^2 - \|\mathbf{E}_{TM}\|^2 d\sigma \right) + \left(\int_{\text{probe2}} \|\mathbf{E}_{TM}\|^2 - \|\mathbf{E}_{TE}\|^2 d\sigma \right)$$

$$\text{subject to } A(\mathbf{x}) \mathbf{E} = \mathbf{b}, \mathbf{x} \in [\mathbf{x}_1, \mathbf{x}_2]$$

提问： 神经网络的结构

神经元的公式 $h_o = f(w_1h_1 + w_2h_2 + w_3h_3 + \dots) = f(\sum_i w_ih_i)$

传递函数公式 $f(x) = s(x) = \frac{1}{1 + e^{-x}}$



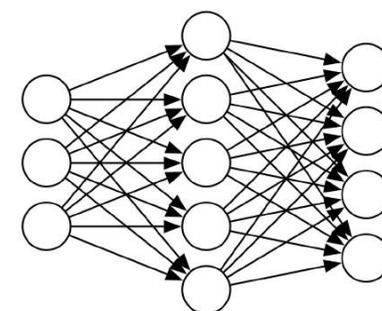
提问： 神经网络的BP算法

逐层求导可得

$$\begin{cases} \mathbf{e}_M = \mathbf{h}_M - \mathbf{y} \\ \mathbf{e}_j = (\mathbf{W}_j^T \mathbf{e}_{j+1}) \circ \mathbf{s}'_{j+1} & (j = 2, \dots, M-1) \\ \frac{\partial \mathbf{E}}{\partial \mathbf{W}_j} = (\mathbf{e}_{j+1} \circ \mathbf{s}'_{j+1}) \mathbf{h}_j^T & (j = 1, \dots, M-1) \end{cases}$$

迭代公式

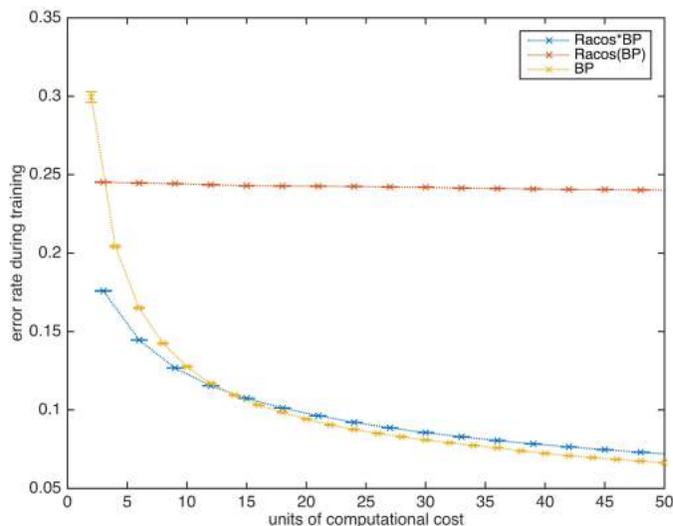
$$\mathbf{W}_i^{(t)} = \mathbf{W}_i^{(t-1)} - \alpha \sum_{(\mathbf{x}, \mathbf{y})} \left(\frac{\partial \mathbf{E}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{W}_i^{(t-1)}} \right)$$



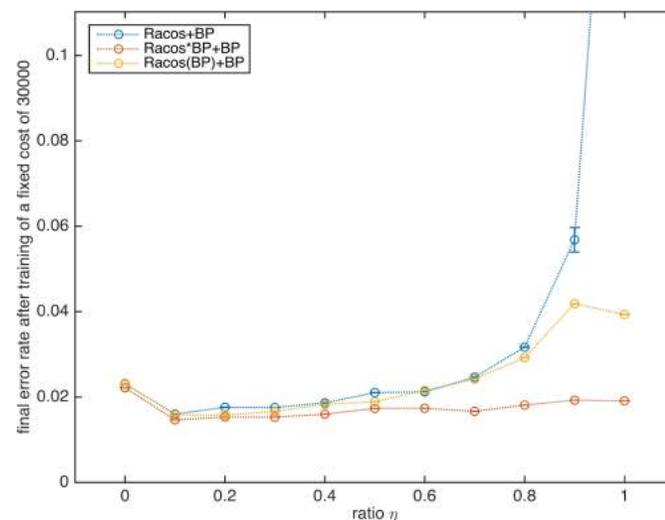
Input Hidden Output

提问： 混合策略训练更详细的结果

Racos在神经网络优化上的应用取得比传统梯度方法更好的效果。



Racos(BP)、*Racos*BP*和*BP*在前50单位计算成本下神经网络误差率的下降趋势。x轴表示计算成本，y轴表示神经网络在MNIST测试数据集上的误差率。在较小的计算成本下，*Racos*BP*能快速找到比*BP*更好的解。小叉表示60次平行测试的平均值，误差棒表示其标准差。



各个算法在固定30000单位计算成本下的误差率与 η 的关系。x轴表示比例 η ，y轴表示*Racos+BP*、*Racos(BP)+BP*和*Racos*BP+BP*在固定30000单位计算成本下的误差率。从 $\eta = 0$ 和 $\eta = 1$ 端点可以看出*Racos*、*Racos(BP)*、*Racos*BP*和*BP*的误差率。圆圈表示60次平行测试的平均值，误差棒表示其标准差。