Introduction to Big Data Analytics

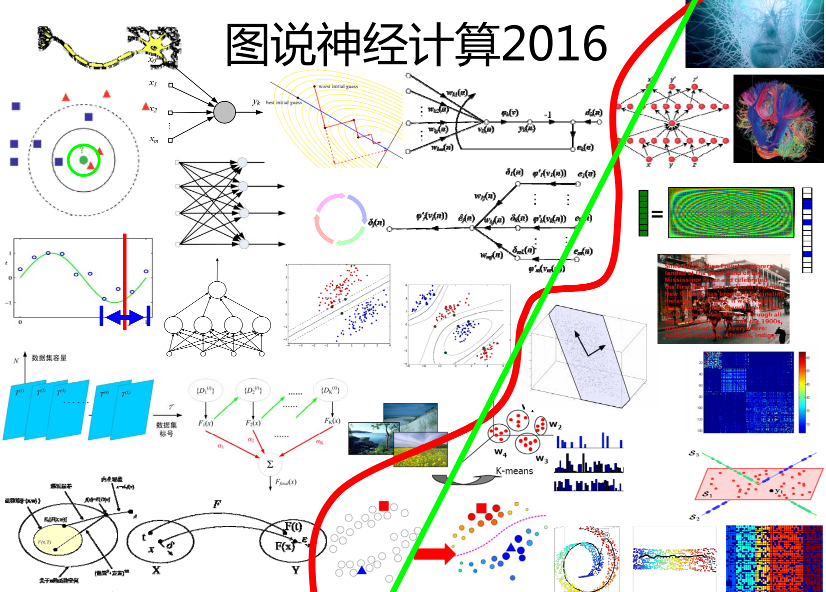
Compressed Sensing and Sparse Representation

Yanwei Fu

School of Data Science, Fudan University

- Highly related to numerical computing, statistical theory
- Very useful in many practical applications
- Basic algorithms for advanced Big data analytics

图说神经计算2016

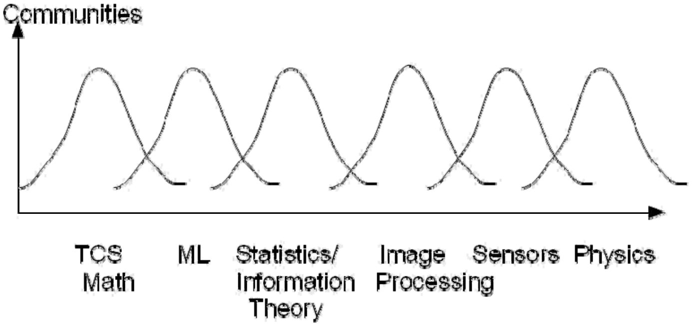

worst initial guess
best initial guess
数据集容量
数据集
标号
K-means
F
X
Y
F(t)
F(x)

内容提要

- 引言
- 压缩感知
- 稀疏表示及其应用
  - SRC / SSC
- 矩阵恢复及其应用
  - LRMC / RPCA(PCP) / LRR
- 最优化问题求解算法
  - MP / OMP / FS / ADMM / (LADMM)

引言

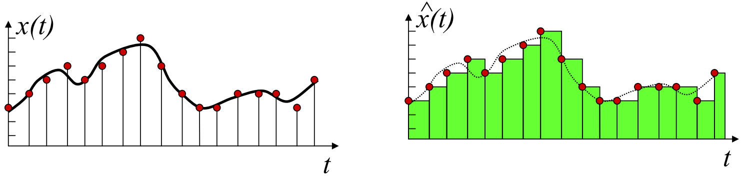
- 压缩感知与稀疏表示
  - **Compressed Sensing & Sparse Representation**



传统数据获取: 采样 (Sampling)

- 采样定理
  - 对于一个限带(**band limited**)信号**x(t)**，准确重构所需要的最小采样速率为

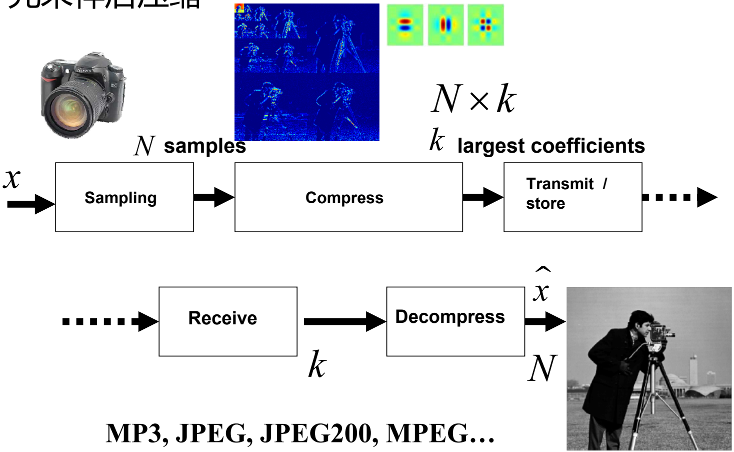
$$f_{sampling} \geq 2 f_{\max}$$



$$x[n] = x(nT),\ n \in Z \qquad x(t) = \sum_{n\in Z} x[n]\, sinc\left(\frac{t}{T} - n\right)$$

传统的数据压缩模式

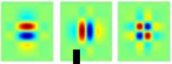
- 先采样后压缩



MP3, JPEG, JPEG200, MPEG…

稀疏信号(Sparse Signal)

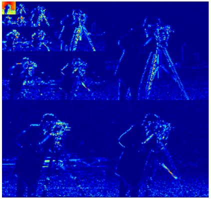
basis functions



$$x \approx \sum_{i=1}^{k} \alpha_i \psi_i$$

$$\alpha_i = \langle x, \psi_i \rangle$$

transform coefficients



k largest coefficients

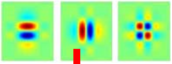
$$k << N$$

- 数字信号往往通常非常稀疏
  - 音频(Audio): MP3, AAC… ~10:1压缩率
  - 图像(Images): JPEG, JPEG2000… ~20:1 压缩率
  - 视频序列(Video sequences): MPEG2, MPEG4… ~40:1 压缩率

k-稀疏信号(k-Sparse Signal)

- 

basis functions



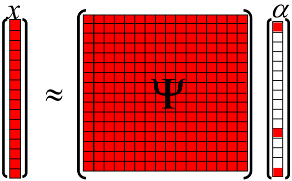
$$x \approx \sum_{i=1}^{k} \alpha_i \psi_i$$

transform coefficients

$$\alpha_i = \langle x, \psi_i \rangle$$

$$k << N$$

x: k-sparse signal in *non-sparse* domain



α : k -sparse signal

k nonzero entries

内容提要

- 引言
- 压缩感知
- 稀疏表示及其应用
  - SRC / SSC
- 矩阵恢复及其应用
  - LRMC / RPCA(PCP) / LRR
- 最优化问题求解算法
  - MP / OMP / FS / ADMM / (LADMM)

压缩感知框架

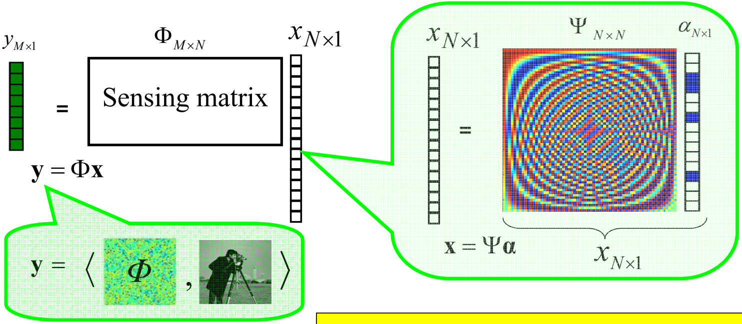
- 编码(Encoding):
  - 获得M个测量系数y，即 $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\alpha}$
    - 其中 $\boldsymbol{\Phi}$ 称为感知矩阵， $\boldsymbol{\Psi}$ 为稀疏化基或稀疏化变换
- 解码(Decoding):
  - 从测量向量y 借助稀疏先验(sparse prior)和非线性优化重构出待观测向量x
    - 求解α

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \mathbf{y} = \Phi\Psi\alpha$$

    - 计算x: $\hat{x} = \Psi\hat{\alpha}$

压缩感知框架 — 编码过程

- 借助感知矩阵 **Φ** 获得被观测向量x的一个M维测量，其中x在变换 **Ψ** 下是稀疏的

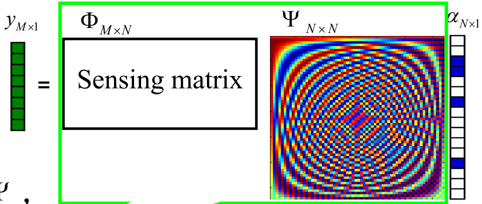


Φ 与 **Ψ** 是不相干的(incoherent)

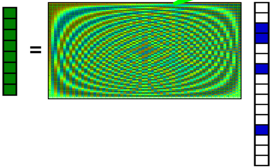
$$\mathbf{y}=\Phi\Psi\boldsymbol{\alpha}=\mathbf{A}\boldsymbol{\alpha}$$

压缩感知框架 — 编码过程

- $\mathbf{y} = \Phi\Psi\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\alpha}$



  – 其中$\mathbf{A} = \Phi\Psi$，
    通常被称之为过完备词典(Over-complete Dictionary)



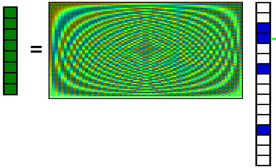
$\mathbf{y} = \mathbf{A} \quad \boldsymbol{\alpha}$

压缩感知框架 — 解码过程

- 给定过完备词典A和测量y，重构出观测向量x
  - 核心问题：如何有效和可靠地从 $\mathbf{y}=\mathbf{A}\boldsymbol{\alpha}$ 恢复出稀疏表示系数 $\boldsymbol{\alpha}$
  - 等价于求解欠定方程组

若无稀疏性先验(sparseness prior): underdetermined (ill-posed) 问题

$\mathbf{y} \;=\; \mathbf{A} \;\; \boldsymbol{\alpha}$



若有稀疏性先验(sparseness prior): well-posed问题

压缩感知框架 — 解码过程

- 稀疏表示系数恢复，即下列优化问题：

P$_0$: $$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\alpha$$

Computationally intractable!
A NP-hard Problem.

P$_1$: $$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\alpha$$

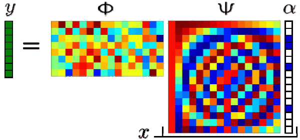
Computationally tractable!
A Convex Problem

基的不相干性定义

- 不相干基(Incoherence bases)
  - 假设信号x在正交规范变换(Orthonormal transform)Ψ下是稀疏的，即

    $\mathbf{x} = \Psi \alpha$ 其中 α 是k-稀疏的

  - 借助感知矩阵Φ获取M个测量，



$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \alpha = \mathbf{A} \alpha$$

其中y是M维列向量， A=Φ Ψ

则Ψ与Φ之间的相干性(coherence)定义为

$$\mu(A) = \mu(\Phi, \Psi) = \sqrt{N} \max_{1 \le i, j \le N} \left| \left\langle \varphi_i^T, \psi_j \right\rangle \right|$$

其中 φ_i : rows of Φ ψ_j : columns of Ψ

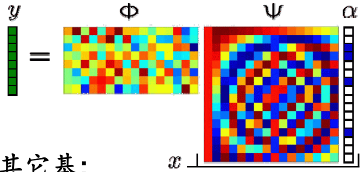
不相干基的性质

- 相干性(coherence)的范围: $1 \le \mu(\Phi, \Psi) \le \sqrt{N}$
  - 如果μ比较小，则称基是不相干的(**Incoherent**)
- 直观解释:
  - 当两个基是不相干的，则**A**中的元素分布得比较分散(**spread out**)，其中**A=ΦΨ**
  - 借助感知矩阵Φ获取**m**个测量，**M**个测量中的每个测量只包含信号**x**的少量信息
    - 我们希望相干性小
- 不相干基举例:
  - **DFT** 和单位阵**I:**

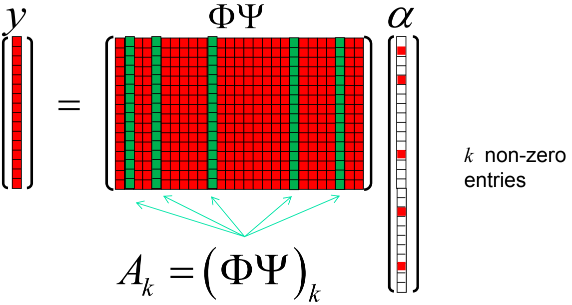
$$\mu = 1$$

  - 高斯或**Bernoulli**矩阵和任何其它基:

$$\mu = \sqrt{2\log N}$$



RIP (Restricted Isometry Property)



- 准确信号恢复的充分条件:

$$(1-\varepsilon)\|\boldsymbol{\alpha}_k\|_2^2 \leq \|A_k\boldsymbol{\alpha}_k\|_2^2 \leq (1+\varepsilon)\|\boldsymbol{\alpha}_k\|_2^2$$

所有由A的k列构成的子矩阵近似正交

准确信号恢复所需的测量数目

- $$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1, \quad \text{s. t.} \quad \mathbf{y} = A\boldsymbol{\alpha} \qquad (1)$$

- 定理(Candes, Romberg & Tao, 2006)

  假设α的**support**为**T**，矩阵**A**的**m**个行向量是从**NxN**的**DFT**矩阵的**N**行中均匀地随机选出，那么如果**m**满足条件**:**

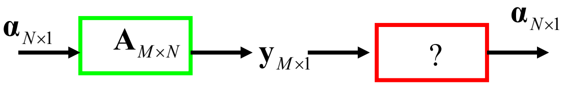
  $$m \geq C|T|\log N$$

  则求解**L1**最小化问题**(1)** 将以概率至少$1-O(N^{-r})$准确地恢复出α

实际上, m为4-6 倍的$|T|$即可保证准确发现

从线性方程组求解看信号重构

- 考虑一个线性信号系统



$$\mathbf{y}_{M\times1} = \mathbf{A}_{M\times N}\boldsymbol{\alpha}_{N\times1}$$

- M = N
  - 方程数目等于未知数个数 → **Critical sampling**
    - 若方程满秩，则有唯一解 $\boldsymbol{\alpha} = \mathbf{A}^{-1}\mathbf{y}$
- M > N
  - 方程数目大于未知数个数 → **Over-sampling**
    - 若方程列满秩，则借助伪逆，可以得到方程唯一解 $\boldsymbol{\alpha} = \mathbf{A}^{+}\mathbf{y}$
- M < N
  - 方程数目小于未知数个数 → **Under-sampling**
    - 无穷解
    - 如果解是稀疏的，则可以有唯一解 $\boldsymbol{\alpha}_s = \mathbf{A}_s{}^{+}\mathbf{y}_s$

内容提要

- 引言
- 压缩感知
- 稀疏表示及其应用
  - SRC / SSC
- 矩阵恢复及其应用
  - LRMC / RPCA(PCP) / LRR
- 最优化问题求解算法
  - MP / OMP / FS / ADMM / (LADMM)

稀疏表示

- 给定过完备词典A, 稀疏表示通过求解下述模型完成:
  - 使用L_0范数

Combinatorial problem

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad y = A\alpha$$

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|y - A\alpha\|_2^2 \le \delta$$

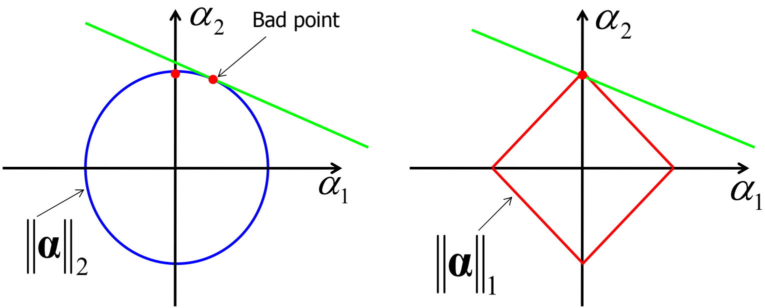
  - 使用L_1范数

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad y = A\alpha$$

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|y - A\alpha\|_2^2 \le \delta$$

Convex problem

Why Is L1 Better Than L2?



The line $\mathbf{y} = \mathbf{A}\boldsymbol{\alpha}$ intersect L_2 circle at a non-sparse point

The line $\mathbf{y} = \mathbf{A}\boldsymbol{\alpha}$ intersect L_1 diamond at the sparse point

Unique and exact solution

稀疏表示的典型应用

- 稀疏表示被广泛应用于信号去噪、恢复、增强中，也被应用于人脸识别、子空间聚类以及特征学习等方面
  – 图像去噪(**De-noising**)
  – 图像增强(**Enhancement**)
  – 光照变化条件下的正面人脸图像识别
  – 子空间聚类
    - 运动分割(Motion Segmentation)
  – 转换学习(**Transfer Learning**)
    - Self-Taught Learning

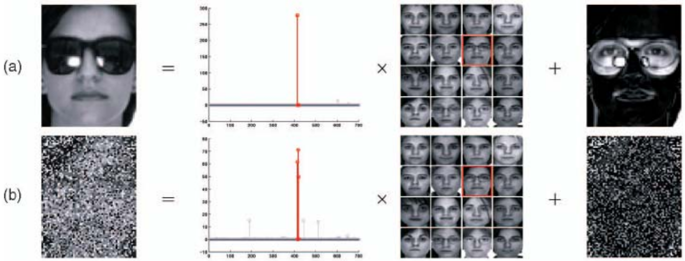
基于稀疏表示的人脸识别

- 把训练数据作为词典A，计算测试数据的稀疏表示，根据稀疏表示的系数进行分类

$$A = [A_1, A_2, \ldots, A_k]$$

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 + \lambda \|\mathbf{y} - A\mathbf{x}\|_2^2$$

$$\min_{\mathbf{x},\mathbf{e}} \|\mathbf{x}\|_1 + \lambda \|\mathbf{e}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x} + \mathbf{e}$$



J. Wright et al.: Robust Face Recognition via Sparse Representation, TPAMI 2009.

SRC 算法

- **Algorithm 1. Sparse Representation-based Classification (SRC)**

  1: **Input:** a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\varepsilon > 0$.)

  2: Normalize the columns of A to have unit ℓ^2-norm.

  3: Solve the ℓ^1-minimization problem:

  $$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad A\boldsymbol{x} = \boldsymbol{y}. \qquad (13)$$

  (Or alternatively, solve

  $$\hat{\boldsymbol{x}}_1 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|A\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \varepsilon.)$$

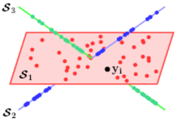
  4: Compute the residuals $r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\,\delta_i(\hat{\boldsymbol{x}}_1)\|_2$ for $i = 1, \ldots, k$.

  5: **Output:** $\text{identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$.

J. Wright et al.: Robust Face Recognition via Sparse Representation, TPAMI 2009.

稀疏子空间聚类(SSC)

- 子空间聚类(Subspace Clustering)
  - 给定一组分布在k个子空间上的数据点，把数据点分成k个子空间

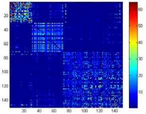


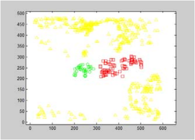
- 以数据矩阵Y本身作为词典(dictionary)，寻找最稀疏的自表示(Self-Expression)

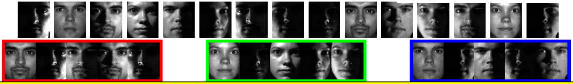
$$\min_{C} \|C\|_0 \quad \text{s.t.} \quad Y = YC, \quad \operatorname{diag}(C) = 0$$

$$\min_{C} \|C\|_1 \quad \text{s.t.} \quad Y = YC, \quad \operatorname{diag}(C) = 0$$

- 应用: (1) 运动分割；
  (2) 人脸图像聚类







Ehsan Elhamifar & Rene Vidal: Sparse Subspace Clustering: Algorithm, Theory, and Applications, TPAMI 2013.

SSC算法

Algorithm 1 : Sparse Subspace Clustering (SSC)

Input: A set of points $\{\boldsymbol{y}_i\}_{i=1}^N$ lying in a union of n linear subspaces $\{\mathcal{S}_i\}_{i=1}^n$.

1: Solve the sparse optimization program:

$$\min \|\boldsymbol{C}\|_1 \quad \text{s.t.} \quad \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \quad \operatorname{diag}(\boldsymbol{C}) = \boldsymbol{0}.$$

2: Form a similarity graph by connecting node i, representing $\boldsymbol{y}_i$, to node j, representing $\boldsymbol{y}_j$, by an edge whose weight is equal to $w_{ij} = |c_{ij}| + |c_{ji}|$.

3: Infer the segmentation of the data from the $n+1$ eigenvectors of the symmetric normalized Laplacian matrix of the graph corresponding to its $n+1$ smallest eigenvalues using the K-means algorithm.

Output: Segmentation of the data: $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$.

$$\min_{C,E} \|C\|_1 + \lambda \|E\|_F \quad \text{s.t.} \quad Y = YC + E, \quad \operatorname{diag}(C) = 0$$

$$\min_{C,E} \|C\|_1 + \lambda \|Y - YC - E\|_F + \gamma \|E\|_1 \quad \text{s.t.} \quad \operatorname{diag}(C) = 0$$

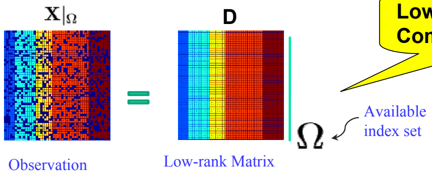
Ehsan Elhamifar & Rene Vidal: Sparse Subspace Clustering: Algorithm, Theory, and Applications, TPAMI 2013.

内容提要

- 引言
- 压缩感知
- 稀疏表示及其应用
  - SRC / SSC
- 矩阵恢复及其应用
  - LRMC / RPCA(PCP) / LRR
- 最优化问题求解算法
  - MP / OMP / FS / ADMM / (LADMM)

矩阵填充(Matrix Completion)

- 动机:
  - 观测到部分数据
  - 把缺失数据准确填充，获得准确的完整数据**X**



Low-Rank Matrix Completion: LRMC

- 基本假设:
  - 数据**X**是低秩的(**Low-Rank**)，观测到的元素是随机均匀分布的
  - 矩阵**X**满足一定的不相干条件

- 基本模型: $\min_{D} \ \operatorname{rank}(D) \ \text{s.t.} \ P_{\Omega}(D) = P_{\Omega}(X)$

$$\boxed{\min_{D} \|D\|_{*} \ \text{s.t.} \ P_{\Omega}(D) = P_{\Omega}(X)}$$

E. J. Candès and B. Recht: Exact matrix completion via convex optimization, 2009.
E. Candès and T. Tao: The power of convex relaxation: Near-optimal matrix completion, 2009.

低秩矩阵填充(LRMC)

- 最优化问题:

$$\min_{D} \|D\|_* \quad \text{s.t. } P_\Omega(D) = P_\Omega(X)$$

- 理论保证:
  - 如果矩阵**X**满足一定的不相干(**Incoherence**)条件，观测的元素均匀分布，则当观测的元素数**k**满足下述条件

$$k > C \cdot r \cdot m \log^2 m$$

  则以极高的成功概率保证求解该凸优化问题能够得到准确的矩阵**X**

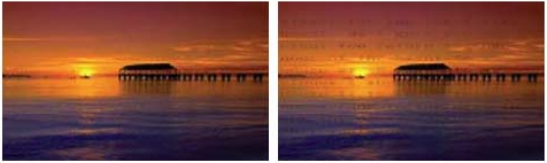
- 如何求解:

Benjamin Recht, "A Simpler Approach to Matrix Completion", Journal of Machine Learning Research Vol. 12, (2011), pp. 3413-3430.

LRMC应用举例: Image impainting

-

(a) original image (b) masked image



Debin Zhang et al. Matrix completion by Truncated Nuclear Norm Regularization, CVPR 2012.

LRMC应用举例: Image impainting

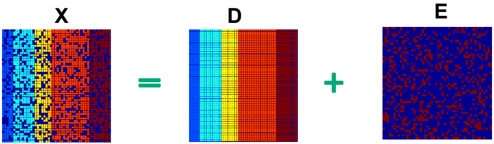
- 

Debin Zhang et al. Matrix completion by Truncated Nuclear Norm Regularization, CVPR 2012.

RPCA: Robust PCA

Wright et al. NIPS2009

- 动机:
  - 从被噪声污染的数据X中恢复出干净数据D_0，分离出噪声污染E_0
    - X = D + E



- 基本假设:
  - 噪声E_0是稀疏的，数据D_0是低秩的，矩阵D_0满足一定的不相干条件

- 基本模型:

$$\min_{D,E} \ \operatorname{rank}(D) + \lambda \|E\|_0 \quad \text{s.t.} \quad X = D + E$$

$$\boxed{\min_{D,E} \|D\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad X = D + E}$$

RPCA: Robust PCA

PCP: Principal Component Pursuit:

- 优化问题:

$$\min_{D,E} \|D\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad X = D + E$$

- 理论保证:
  - 在满足一定条件下，以高概率保证通过求解上述凸优化问题能够恢复出准确的数据矩阵**D_0**和噪声干扰**E_0**，其中

$$\lambda = 1/\sqrt{m}$$

D **是大小为** $n_1 \times n_2$ **的矩阵** $m = \max\{n_1, n_2\}$

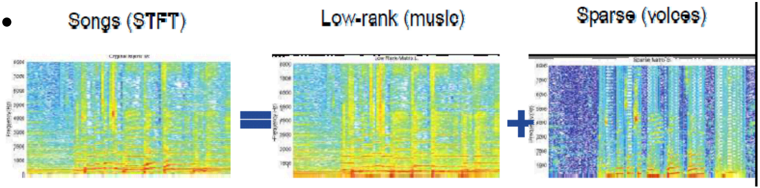
Emmanual J. Candes, Xiaolong Li, Yi Ma, and John Wright: Robust Principal Component Analysis? Journal of the ACM, 2011.

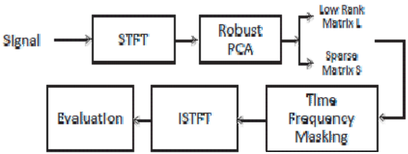
RPCA应用: 视频中背景与目标分离

- $X \quad = \quad D \quad + \quad E$

(a) (b) (c)

RPCA应用: 歌声与背景音乐的分离

- 



Min, Zhang, Wright, Ma, CIKM 2010 / Sprechmann, Bronstein, Sapiro, ISMIR 2012

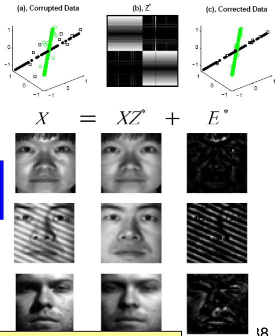
低秩表示(LRR)

- Low-Rank Representation
  - 以数据矩阵X本身作为词典(dictionary) A, 寻找最低秩的自表示(Self-Expression)

$$\min_{Z} \operatorname{rank}(Z) \quad \text{s.t.} \quad X = XZ$$

$$\boxed{\min_{Z} \|Z\|_* \quad \text{s.t.} \quad X = XZ}$$

$$\boxed{\min_{Z,E} \|Z\|_* + \lambda \|E\|_{2,1} \quad \text{s.t.} \quad X = XZ + E}$$



G.Liu et al.: Robust Subspace Segmentation by Low-Rank Representation, ICML2010, TPAMI 2013.

内容提要

- 引言
- 压缩感知
- 稀疏表示及其应用
  - SRC / SSC
- 矩阵恢复及其应用
  - LRMC / RPCA(PCP) / LRR
- 目标函数不可微的带约束最优化问题求解
  - MP / OMP / FS / ISTA / FISTA /ADMM / (LADMM)

最优化问题

- 形式上的区别
  - 目标函数可导且无约束
  - 目标函数可导但有约束
  - 目标函数不可导但无约束
    - 比如: L1范数，核(Nuclear)范数
  - 目标函数不可导且有约束

- 本质上的区别
  - 凸(convex)优化问题
  - 非凸(nonconvex)优化问题

L_0-最小化问题

- 以L_0范数为目标函数，寻找最稀疏表示

$$\boxed{\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad y = A\alpha}$$

Combinatorial problem

  - 若考虑到高斯噪声，则：

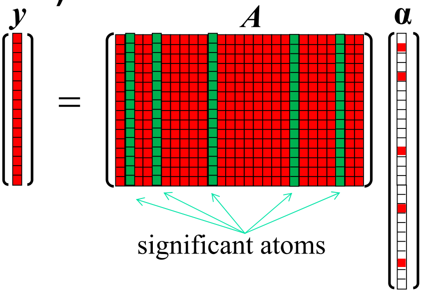
$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|y - A\alpha\|_2^2 \le \delta$$

  - 若考虑到稀疏噪声，则：

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|y - A\alpha\|_1 \le \delta$$

L_0-最小化问题的求解

- 贪婪(Greedy)算法
  - 匹配追踪(**Matching Pursuit: MP)**
  - 正交匹配追踪(**Orthogonal Matching Pursuit: OMP)**



在每次迭代过程中，尝试确认出字典A中与非零表示系数相对应的列

匹配追踪(MP)

- 基本思想:
  - 每次迭代时，尝试确认出最突出的基向量
- 算法步骤:
  - **1.** 令**t=1**，设置残差向量 $\mathbf{r}_t = \mathbf{y}$ 和下标集 $I_t = \varnothing$
  - **2.** 寻找与残差相关度最大的基向量下标
    $j = \arg\max_{1 \le i \le N} \left| \langle \mathbf{r}_t, \mathbf{a}_i \rangle \right|$，加入下标集 $I_t \leftarrow I_t \cup \{j\}$
  - **3.** 更新残差 $\mathbf{r}_{t+1} = \mathbf{r}_t - \langle \mathbf{r}_t, \mathbf{a}_j \rangle \mathbf{a}_j$
  - **4.** 若**t=k**, 则停止; 否则, 令**t←t+1**, 转第**2**步.

S. Mallat and Z. Zhang: Matching pursuit in a time-frequency dictionary, IEEE Trans. Signal Proc., 41 (1993), pp. 3397–3415.

正交匹配追踪(OMP)

- 基本思想:
  - 保证残差与此前所选择的所有基向量正交
- 算法步骤:
  - 1. 令t=1，设置残差向量 $\mathbf{r}_t = \mathbf{y}$ 和下标集 $I_t = \varnothing$
  - 2. 寻找与残差相关度最大的基向量下标 $j = \arg\max_{1 \le i \le N} \left| \langle \mathbf{r}_t, \mathbf{a}_i \rangle \right|$，加入下标集 $I_t \leftarrow I_t \cup \{j\}$
  - 3. 计算表达系数 $\mathbf{c}_t^* = \arg\min_{\mathbf{c}} \left\| \mathbf{y} - A_{I_t} \mathbf{c} \right\|_2^2$
  - 5. 更新残差 $\mathbf{r}_{t+1} = \mathbf{y} - A_{I_t} \mathbf{c}_t^*$
  - 6. 若t=k, 则停止; 否则, 令t←t+1, 转第2步.

Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in 27th Asilomar Conf. on Signals, Systems and Comput., Nov. 1993.

L_1-最小化问题

- 把L_0范数放松为L_1范数

Combinatorial problem

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad y = A\alpha$$

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad \|y - A\alpha\|_2^2 \le \delta$$

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad y = A\alpha$$

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad \|y - A\alpha\|_2^2 \le \delta$$

Convex problem

L_1-最小化问题的求解

- 基追踪(Basis Pursuit: BP)
  - 使用变量分裂法转化为线性规划**(Linear Programming)**问题
    - 求解LP问题的方法，比如单纯形(Simplex)法 / 内点法(interior-point)
- 特征符号搜索(Feature Sign Search: FSS)
- 迭代收缩阈值算法(ISTA)
  - 快速迭代收缩阈值算法**(FISTA)**

[1] S. Chen, D. Donoho, and M. Saunders: Atomic Decomposition by Basis Pursuit, SIAM Review, Vol. 43, No. 1, pp. 129-159, 2001.
[2] H. Lee, A. Battle, R. Raina, and Andrew Y. Ng: Efficient Sparse Coding Algorithms, NIPS 2007.

基追踪 (Basis Pursuit)

- 考虑优化问题 $\min \|\alpha\|_1 \quad \text{s.t.} \quad y = A\alpha$

令 $\alpha = u - v \quad \text{with} \quad u \geq 0, v \geq 0$

则 $\|\alpha\|_1 = 1^T u + 1^T v$ 令 $z \equiv \begin{bmatrix} u \\ v \end{bmatrix}; \quad B = \begin{bmatrix} A & -A \end{bmatrix}$

从而

$$\boxed{\begin{array}{lll} \min\limits_{\alpha} \|\alpha\|_1 & \Rightarrow & \min\limits_{z} \; 1^T z \\ \text{s.t.} \quad y = A\alpha & & \text{s.t.} \quad Bz = y, \quad z \geq 0 \end{array}}$$

– 线性规划问题的算法：**Simplex / primal-dual interior-point /log-barrier…**

S. Chen, D. Donoho, and M. Saunders: Atomic Decomposition by Basis Pursuit, SIAM Review, Vol. 43, No. 1, pp. 129-159, 2001.

L_1-最小化问题的闭式解

- 考虑一个特殊的L_1-最小化问题 (A=I)

$$\min_{\mathbf{x}} \lambda \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}-\mathbf{c}\|_2^2$$

其最优解为： $\mathbf{x}^* = T_\lambda(\mathbf{c})$

– 其中 $x_j^* = T_\lambda(c_j) = \begin{cases} c_j - \lambda, \text{ if } & c_j > \lambda \\ 0, \text{ if } & |c_j| \le \lambda \\ c_j + \lambda \text{ if } & c_j < -\lambda \end{cases}$

Soft-Thresholding 算子

特征符号搜索 (Feature Sign Search)

- Feature Sign Search 算法基本思路
  - 考虑下述**L1**最小化问题

$$\arg\min_{\mathbf{x}} \|A\mathbf{x}-\mathbf{y}\|_2^2 + \gamma\|\mathbf{x}\|_1$$

  - 注意到问题的最优性条件为 $\dfrac{\partial\|A\mathbf{x}-\mathbf{y}\|_2^2}{\partial\mathbf{x}} + \gamma\cdot\partial\|\mathbf{x}\|_1 = 0$

    - 对于非零系数，有

$$\frac{\partial\|A\mathbf{x}-\mathbf{y}\|_2^2}{\partial x_j} + \gamma\,\mathrm{sign}(x_j) = 0$$

    - 对于零系数 ，有

$$\left|\frac{\partial\|A\mathbf{x}-\mathbf{y}\|_2^2}{\partial x_j}\right| \le \gamma$$

  - 如果**x**的分量的符号已知，则**L1**范数可被化简，相应的子问题转化为无约束二次规划**(QP)**问题

Feature Sign Search 算法

Algorithm 1 Feature-sign search algorithm

1 Initialize $x := \vec{0}$, $\theta := \vec{0}$, and *active set* $:= \{\}$, where $\theta_i \in \{-1, 0, 1\}$ denotes $\text{sign}(x_i)$.

2 From zero coefficients of x, select $i = \arg\max_i \left|\frac{\partial\|y-Ax\|^2}{\partial x_i}\right|$.

Activate x_i (add i to the *active set*) only if it locally improves the objective, namely:

If $\frac{\partial\|y-Ax\|^2}{\partial x_i} > \gamma$, then set $\theta_i := -1$, *active set* $:= \{i\} \cup$ *active set*.

If $\frac{\partial\|y-Ax\|^2}{\partial x_i} < -\gamma$, then set $\theta_i := 1$, *active set* $:= \{i\} \cup$ *active set*.

3 Feature-sign step:

Let $\hat{A}$ be a submatrix of A that contains only the columns corresponding to the *active set*.

Let $\hat{x}$ and $\hat{\theta}$ be subvectors of x and θ corresponding to the *active set*.

Compute the analytical solution to the resulting unconstrained QP ($\text{minimize}_{\hat{x}} \|y - \hat{A}\hat{x}\|^2 + \gamma\hat{\theta}^\top\hat{x}$):

$\hat{x}_{new} := (\hat{A}^\top\hat{A})^{-1}(\hat{A}^\top y - \gamma\hat{\theta}/2)$,

Perform a discrete line search on the closed line segment from $\hat{x}$ to $\hat{x}_{new}$:

Check the objective value at $\hat{x}_{new}$ and all points where any coefficient changes sign.

Update $\hat{x}$ (and the corresponding entries in x) to the point with the lowest objective value.

Remove zero coefficients of $\hat{x}$ from the *active set* and update $\theta := \text{sign}(x)$.

4 Check the optimality conditions:

(a) Optimality condition for nonzero coefficients: $\frac{\partial\|y-Ax\|^2}{\partial x_j} + \gamma\,\text{sign}(x_j) = 0, \forall x_j \neq 0$

If condition (a) is not satisfied, go to Step 3 (without any new activation); else check condition (b).

(b) Optimality condition for zero coefficients: $\left|\frac{\partial\|y-Ax\|^2}{\partial x_j}\right| \leq \gamma, \forall x_j = 0$

If condition (b) is not satisfied, go to Step 2; otherwise return x as the solution.

H. Lee, A. Battle, R. Raina, and Andrew Y. Ng: Efficient Sparse Coding Algorithms, NIPS2007.

迭代收缩阈值算法(ISTA)

- Iterative Shrinkage Thresholding Algorithm (ISTA)
  - 考虑**L_1-**最小化问题 $\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|A\mathbf{x}-\mathbf{c}\|_2^2$
    - 令 $f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}-\mathbf{c}\|_2^2$
    - 在x_k处，对f(x)进行2阶泰勒级数展开, 其中L为Lipschitz常数:

$$f(\mathbf{x}) \le Q(\mathbf{x},\mathbf{x}_k) = f(\mathbf{x}_k) + \left\langle \mathbf{x}-\mathbf{x}_k, \nabla f(\mathbf{x}_k)\right\rangle + \frac{L}{2}\|\mathbf{x}-\mathbf{x}_k\|_2^2$$

    - 则 $\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + Q(\mathbf{x},\mathbf{x}_k)$

$$= \arg\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + \frac{L}{2}\left\|\mathbf{x} - \left(\mathbf{x}_k - \frac{1}{L}\nabla f(\mathbf{x}_k)\right)\right\|^2$$

  - **ISTA算法:**
    - 令k=0，初始化x_0; 若不满足收敛条件，计算x_{k+1}，并令$k \leftarrow k+1$

快速迭代收缩阈值算法(FISTA)

- Fast Iterative Shrinkage Thresholding Algorithm (FISTA)
  - 考虑L_1-最小化问题 $\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + f(\mathbf{x})$
    - 初始化: $y_k = x_{k-1}$, $t_k=1$, $k=1$
    - 在y_k处， 对f(x)进行2阶泰勒级数展开, 其中L为Lipschitz常数:

$$f(\mathbf{x}) \le Q(\mathbf{x}, \mathbf{y}_k) = f(\mathbf{y}_k) + \langle \mathbf{x} - \mathbf{y}_k, \nabla f(\mathbf{y}_k) \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{y}_k\|_2^2$$

    - 则 $\mathbf{x}_k = \arg\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1 + \frac{L}{2}\left\|\mathbf{x} - \left(\mathbf{y}_k - \frac{1}{L}\nabla f(\mathbf{y}_k)\right)\right\|^2$

$$t_{k+1} = \left(1 + \sqrt{1 + 4t_k^2}\right)/2, \qquad \mathbf{y}_{k+1} = \mathbf{x}_k + (t_k - 1)(\mathbf{x}_k - \mathbf{x}_{k-1})/t_{k+1}$$

FISTA算法:

- 初始化x_0, 令 $y_k = x_{k-1}$, $t_k=1$, $k=1$; 若不满足收敛条件, 不断计算 x_k , t_{k+1} , y_{k+1} ,并令$k \leftarrow k+1$

[1] Beck, A; Teboulle, M (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". SIAM J. Imaging Science, pp. 183–202.

秩最小化问题

- 矩阵的最佳低秩近似问题:

$$\min_{D} \|D - X\|_F^2 \quad \text{s.t.} \quad \mathrm{rank}(D) \le r$$

  - 对矩阵X进行奇异值分解**(Singular Value Decomposition: SVD)**，得出:

$$X = USV^T$$

  - 那么，矩阵X的最佳低秩近似为 $D^* = U_r S_r V_r^T$
    其中 $S_r = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_r), \sigma_1 \ge \sigma_2 \ge ... \ge \sigma_r \ge 0$

秩最小化问题

- 矩阵填充问题:

$$\min_{D}\ \operatorname{rank}(D)\ \text{ s.t. }\ P_{\Omega}(D)=P_{\Omega}(X)$$

放松为:

$$\min_{D}\|D\|_{*}\ \text{ s.t. }\ P_{\Omega}(D)=P_{\Omega}(X)$$

  – **奇异值阈值化(Singular Value Thresholding: SVT)算法**

[1] J.-F. Cai, E. J. Cand´es, Z. Shen: A singular value thresholding algorithm for matrix completion, SIAM Journal of Optimization 20 (4) (2008) 1956–1982.

SVT算子

- 最优化问题

$$\min_{D} \lambda \|D\|_* + \frac{1}{2}\|D - X\|_F^2$$

  - 其中*表示矩阵的核(Nuclear)范数, 等于D的奇异值的和

其最优解为

$$D^* = \arg\min_{D} \lambda \|D\|_* + \frac{1}{2}\|D - X\|_F^2 = UT_\lambda(S)V^T$$

  - 其中$[U,S,V] = \mathrm{svd}(X)$为矩阵X的奇异值分解(SVD: Singular Value Decomposition)
  - $T_\lambda(S)$ 表示对矩阵对角线上的元素进行Soft Thresholding处理

[1] J.-F. Cai, E. J. Cand´ es, Z. Shen: A singular value thresholding algorithm for matrix completion, SIAM Journal of Optimization 20 (4) (2008) 1956–1982.

其它非可微目标函数最小化问题

- RPCA问题 (核范数+L_1范数)

$$\min_{D,E} \|D\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad X = D + E$$

- SRC问题:

$$\min_{\mathbf{x},\mathbf{e}} \|\mathbf{x}\|_1 + \lambda \|\mathbf{e}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{x} + \mathbf{e}$$

- SSC问题:

$$\min_{C,E,G} \|C\|_1 + \lambda \|E\|_1 + \gamma \|\mathrm{G}\|_\mathrm{F}^2 \quad \text{s.t.} \quad X = XC + E + G, \quad \mathrm{diag}(C) = 0$$

- LRR问题:

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_{2,1} \quad \text{s.t.} \quad X = XZ + E$$

求解一般的目标函数不可导优化问题

- 增广拉格朗日乘子法
  - **Augmented Lagrange Multiplier: ALM**

- 乘子的交替方向法
  - **Alternating Direction Method of Multipliers: ADMM**

[1] Dimitri P. Bertsekas: Constrained Optimization and Lagrange Multiplier Methods, 1982.
[2] S. Boyd, et al. : Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, FTML 2010.
[3] Lin, Chen, Wu, and Ma, "*The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrix*", 2009.

增广拉格朗日乘子法(ALM)

- 考虑如下有约束优化问题

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}$$

  - 建立拉格朗日辅助函数

$$L_0(\mathbf{x}, \Lambda) = f(\mathbf{x}) + \Lambda^T (A\mathbf{x} - \mathbf{b})$$

  - 建立增广拉格朗日辅助函数

$$L_\mu(\mathbf{x}, \Lambda) = f(\mathbf{x}) + \Lambda^T (A\mathbf{x} - \mathbf{b}) + \frac{\mu}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2$$

- 增广拉格朗日乘子法(Augmented Lagrange Method)

$$\boxed{\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} L_\mu(\mathbf{x}, \Lambda^k)}$$

$$\boxed{\Lambda^{k+1} = \Lambda^k + \mu\left(A\mathbf{x}^{k+1} - \mathbf{b}\right)}$$

也叫乘子法(Method of Multipliers)

Dimitri P. Bertsekas: Constrained Optimization and Lagrange Multiplier Methods, 1982

乘子的交替方向法 (ADMM)

ADMM: Alternating Direction Method of Multipliers

- 考虑有约束优化问题

$$\min_{\mathbf{x},\mathbf{z}} f(\mathbf{x})+g(\mathbf{z}) \quad \text{s.t.} \quad A\mathbf{x}+B\mathbf{z}=\mathbf{c}$$

– 建立**Augmented Lagrangian**

$$L_\mu(\mathbf{x},\mathbf{z},\Lambda)=f(\mathbf{x})+g(\mathbf{z})+\Lambda^T(A\mathbf{x}+B\mathbf{z}-\mathbf{c})+\frac{\mu}{2}\|A\mathbf{x}+B\mathbf{z}-\mathbf{c}\|_2^2$$

– **ADMM 算法由下列迭代构成:**

$$\left(\mathbf{x}^{k+1},\mathbf{z}^{k+1}\right)=\arg\min_{\mathbf{x},\mathbf{z}} L_\mu\left(\mathbf{x},\mathbf{z},\Lambda^k\right)$$
$$\Lambda^{k+1}=\Lambda^k+\mu\left(A\mathbf{x}^{k+1}+B\mathbf{z}^{k+1}-\mathbf{c}\right)$$

乘子法(MM: Method of Multipliers)

$$\mathbf{x}^{k+1}=\arg\min_{\mathbf{x}} L_\mu\left(\mathbf{x},\mathbf{z}^{k+1},\Lambda^k\right)$$
$$\mathbf{z}^{k+1}=\arg\min_{\mathbf{z}} L_\mu\left(\mathbf{x}^{k+1},\mathbf{z},\Lambda^k\right)$$
$$\Lambda^{k+1}=\Lambda^k+\mu\left(A\mathbf{x}^{k+1}+B\mathbf{z}^{k+1}-\mathbf{c}\right)$$

[1] Zhouchen Lin et al., Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix, Technical Report, UIUC, August 2009.
[2] S. Boyd, et al. : Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, FTML 2010.

ALM / ADMM应用举例: 1 (1/4)

- 考虑下述L1最小化问题

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- 该问题等价为下述有约束L1最小化问题

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{y}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{y}=\mathbf{x}$$

- 构造增广(Augmented)拉格朗日辅助函数

$$L(\mathbf{x},\mathbf{y},\Lambda) = \|A\mathbf{y}-\mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 + \langle \mathbf{y}-\mathbf{x},\Lambda\rangle \boxed{+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2}$$

  - 求解上述无约束优化问题

ALM / ADMM应用举例: 1 (2/4)

- 求解无约束优化问题

$$L(\mathbf{x},\mathbf{y},\Lambda)=\|A\mathbf{y}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_1+\langle\mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

– 交替求解下列子优化问题，直到收敛：

Primal updating: minimization

同时优化x和y可能存在困难

$$\arg\min_{(\mathbf{x},\mathbf{y})} L(\mathbf{x},\mathbf{y},\Lambda)=\|A\mathbf{y}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_1+\langle\mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

$$\arg\max_{\Lambda} L(\mathbf{x},\mathbf{y},\Lambda)=\langle\mathbf{y}-\mathbf{x},\Lambda\rangle$$

Dual updating: maximization

ALM / ADMM应用举例: 1 (3/4)

- 求解无约束优化问题

$$L(\mathbf{x},\mathbf{y},\Lambda)=\|A\mathbf{y}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_1+\langle \mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

– 交替求解下列容易求解的子问题, 直到收敛:

Primal updating: minimization

$$\arg\min_{\mathbf{x}} L(\mathbf{x},\mathbf{y},\Lambda)=\lambda\|\mathbf{x}\|_1+\langle \mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

$$\arg\min_{\mathbf{y}} L(\mathbf{x},\mathbf{y},\Lambda)=\|A\mathbf{y}-\mathbf{b}\|_2^2+\langle \mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

$$\arg\max_{\Lambda} L(\mathbf{x},\mathbf{y},\Lambda)=\langle \mathbf{y}-\mathbf{x},\Lambda\rangle$$

Dual updating: maximization

ALM / ADMM应用举例: 1 (4/4)

- 求解无约束优化问题

$$L(\mathbf{x},\mathbf{y},\Lambda)=\|A\mathbf{y}-\mathbf{b}\|_2^2+\lambda\|\mathbf{x}\|_1+\langle\mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2$$

  – 交替求解下列容易求解的子问题, 直到收敛:

Primal updating: minimization

$$\arg\min_{\mathbf{x}}\lambda\|\mathbf{x}\|_1+\frac{\mu}{2}\|\mathbf{x}-\mathbf{y}-\Lambda/\mu\|_2^2 \qquad =T_{\lambda/\mu}(\mathbf{y}_{k+1}+\Lambda_{k+1}/\mu)$$

$$\arg\min_{\mathbf{y}}\|A\mathbf{y}-\mathbf{b}\|_2^2+\langle\mathbf{y}-\mathbf{x},\Lambda\rangle+\frac{\mu}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 \qquad =(A^TA+\mu\mathrm{I})^{-1}(A^T\mathbf{b}+\mu\mathbf{x}-\Lambda)$$

$$\Lambda_{k+1}=\Lambda_k+\mu_{k+1}(\mathbf{y}_{k+1}-\mathbf{x}_{k+1})$$

Dual updating: maximization

ALM / ADMM应用举例: 2 (1/2)

- 考虑矩阵填充问题

$$\min_D \|D\|_* \ \text{ s.t. } P_\Omega(D) = P_\Omega(X)$$

– 构造增广(Augmented)拉格朗日辅助函数

$$L(D,\Lambda) = \|D\|_* \boxed{+\langle \Lambda, P_\Omega(D-X)\rangle + \frac{\mu}{2}\|P_\Omega(D-X)\|_F^2}$$

- 如何更新D？
  - 保留与D有关的项，记为g(D)，对g(D)在Dk处进行Taylor级数展开

$$\begin{aligned} g(D) &= \|D\|_* + \langle \Lambda, P_\Omega(D-X)\rangle + \frac{\mu}{2}\|P_\Omega(D-X)\|_F^2 \\ &\le \|D\|_* + g(D_k) + \langle \nabla g(D_k), D-D_k\rangle + \frac{\eta\mu}{2}\|D-D_k\|_F^2 \\ &= h(D) \end{aligned}$$

其中 $\nabla g(D_k) = P_\Omega(\Lambda_k) + \mu P_\Omega(D_k - X),\ \eta = 1$

ALM / ADMM应用举例: 2 (2/2)

- 通过求解h(D)来更新D:

$$D_{k+1} = \arg\min_{D} h(D) = \|D\|_* + g(D_k) + \langle \nabla g(D_k), D - D_k \rangle + \frac{\eta\mu}{2}\|D - D_k\|_F^2$$

– 交替求解下列容易求解的子问题, 直到收敛:

Primal updating: minimization

$$\boxed{D_{k+1} = \arg\min_{D} h(D) = \frac{1}{\eta\mu}\|D\|_* + \frac{1}{2}\|D - D_k + \nabla g(D_k)/\eta\mu\|_F^2}$$

$$= SVT_{1/\eta\mu}\left(D_k - \nabla g(D_k)/\eta\mu\right)$$

$$\boxed{\Lambda_{k+1} = \Lambda_k + \mu_{k+1} P_{\Omega}(D_{k+1} - X)}$$

Dual updating: maximization

扩展阅读推荐

- 压缩感知的简介
  - **Emmanuel Candes and Mechael Walin: An Introduction to Compressive Sampling, 2010.**
  - **David Donoho, "Compressed sensing,” IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, Apr. 2006.**
- 理论论文精要
  - 稀疏编码
    - Emmanuel Candès and Terence Tao, "Decoding by linear programming” IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005
    - Emmanuel Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, Feb. 2006.
  - 矩阵填充
    - E. Cand`es, B. Recht, Exact matrix completion via convex optimization, Foundations of Computational Mathematics 9 (2008) 717–772.
    - E. Cand`es, T. Tao, The power of relaxation: Near-optimal matrix completion, IEEE Transactions on Information Theory 56 (5) (2010) 2053–2080.
    - B. Recht, A simpler approach to matrix completion, Journal of Machine Learning Research 12 (2011) 3413–3430.

扩展阅读推荐

- 优化算法
  - **S. Boyd, et al. : Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundation and Trends in Machine Learning, Vol.3, No.1, pp.1-122, 2010.**
  - **Neal Parikh and Stephen Boyd: Proximal Algorithms, Foundations and Trends in Optimization, Vol. 1, No. 3, pp.1-108, 2013.**

- 压缩感知与稀疏表示的专著
  - **Michael Elad: Sparse and Redundant Representations: From Theory to Application in Signal and Image Processing, 2010.**
  - **Michael Elad et al: Compressed Sensing: Theory and Applications, 2012.**

- 数据建模Beyond PCA:
  - **R. Vidal, Y. Ma, and S. Sastry: Generalized Principal Component Analysis. Springer Verlag, 2015. (In press)**

Proximal Method

- Proximal Operator: $\mathbf{prox}_f : \mathbf{R}^n \to \mathbf{R}^n$ of f
  - **f是一个convex 函数**

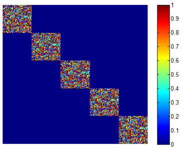
$$\mathbf{prox}_f(v) = \underset{x}{\operatorname{argmin}} \left(f(x) + (1/2)\|x - v\|_2^2 \right)$$

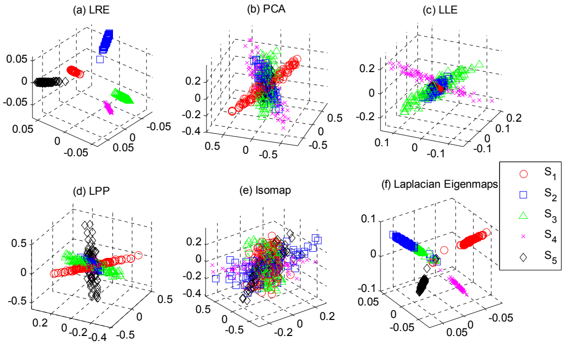
  - **proximal算子是在使用closed-form解去求解一个小的convex优化问题**
- Proximal 算法
  - 用于使用目标函数的**proximal**算子求解一个凸优化问题

Neal Parikh and Stephen Boyd: Proximal Algorithms, Foundations and Trends in Optimization, Vol. 1, No. 3, pp.1-108, 2013.

LRE (Low-Rank Embedding)

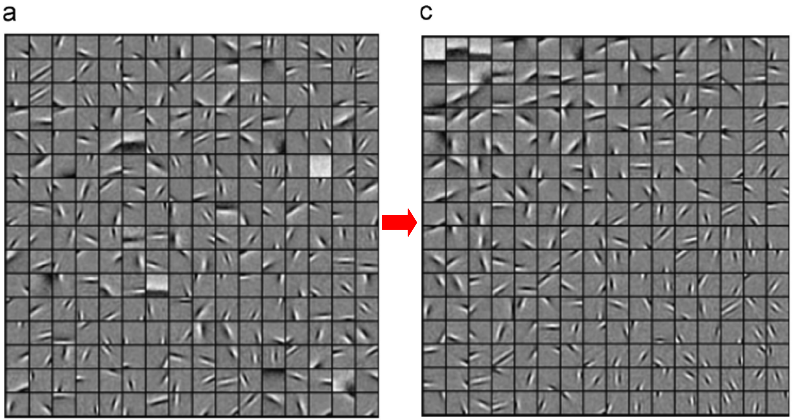
- 低秩嵌入(LRE)降维算法





C.-G. Li et al.: Dimensionality Reduction by Low-Rank Embedding, IScIDE 2012.

基排序: Bases Sorting



[1] Chun-Guang Li, et al. Bases Sorting: Generalizing the Concept of Frequency for Over-complete Dictionaries, Neurocomputing, Vol.115, Sept. 4, 2013, pp.192–200.

结构化的稀疏子空间聚类(S^{3}C)

- 结构化SSC可以同时学习表示矩阵C和子空间分割矩阵Q:

$$\min_{C,E,Q} \|C\|_{1,Q} + \lambda \|E\|_l$$

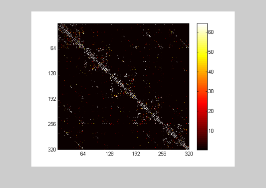
$$\text{s.t. } X = XC + E, \ \operatorname{diag}(C) = 0, \ Q \in \Gamma$$

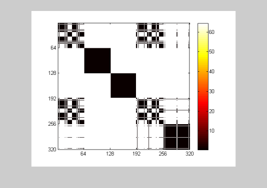
  – **其中结构化L$_1$ 范数定义如下:**

$$\|C\|_{1,Q} = \left\| \left(\alpha\Theta + 11^T \right) \odot C \right\|_1$$

Q为子空间分割矩阵

$$\Theta_{ij} = \frac{1}{2} \left\| \mathbf{q}^{(i)} - \mathbf{q}^{(j)} \right\|_2^2, \qquad Q_{N\times k} = \begin{pmatrix} \mathbf{q}^{(1)} \\ \mathbf{q}^{(2)} \\ \cdots \\ \mathbf{q}^{(N)} \end{pmatrix}$$





[1] Chun-Guang Li and René Vidal, “Structured Sparse Subspace Clustering: A Unified Optimization Framework”, In: IEEE CVPR 2015.

可扩展(Scalable)的子空间聚类算法

- 基于OMP的稀疏子空间聚类
  - 使用**OMP**快速求解**SSC**中的稀疏表示系数计算问题
  - 并给出最优解正确性的理论保证

[1] Chong You and R. Vidal, "Sparse Subspace Clustering by Orthogonal Matching Pursuit", CVPR 2016.

- 基于弹性网络的可扩展子空间聚类算法
  - **L1+L2**
  - 基于主动集的快速求解算法
  - 解的正确性的理论保证
  - 连通性的分析

[2] Chong You, Chun-Guang Li, Daniel Robinson, and Rene Vidal, "Oracle Based Active Set Algorithm for Scalable Elastic Net Subspace Clustering", CVPR 2016.

These slides are adopted from Dr. Chunguang Li.