BCD & ADMM 算法收敛理论

Lecture 10: ADMM 变形技巧与应用举例

罗自炎

北京交通大学数统学院

E-mail: zyluo@bjtu.edu.cn

参考资料

- 教材与参考文献：
  - 最优化：建模、算法与理论
  - Bolte et al., MP 2014
  - Fazel et al., SIMAX 2013
  - Rockafellar & Wets, Variational Analysis
- 致谢： 北京大学文再文教授; 清华大学张立平教授

回顾: 交替方向乘子法ADMM

▶典型优化问题形式:

$$\begin{aligned} \min_{x_1,x_2} \quad & f_1(x_1) + f_2(x_2), \\ \textbf{s.t.} \quad & A_1x_1 + A_2x_2 = b. \end{aligned} \tag{1}$$

▶典型优化问题(1)的增广拉格朗日函数:

$$\begin{aligned} L_\rho(x_1, x_2, y) = & f_1(x_1) + f_2(x_2) + y^{\mathrm{T}}(A_1x_1 + A_2x_2 - b) \\ & + \frac{\rho}{2}\|A_1x_1 + A_2x_2 - b\|_2^2. \end{aligned} \tag{2}$$

▶ ALM迭代:

$$(x_1^{k+1}, x_2^{k+1}) = \arg\min_{x_1,x_2} L_\rho(x_1, x_2, y^k), \tag{3}$$

$$y^{k+1} = y^k + \tau\rho(A_1x_1^{k+1} + A_2x_2^{k+1} - b). \tag{4}$$

▶ ADMM迭代:

$$x_1^{k+1} = \arg\min_{x_1} L_\rho(x_1, x_2^k, y^k), \tag{5}$$

$$x_2^{k+1} = \arg\min_{x_2} L_\rho(x_1^{k+1}, x_2, y^k), \tag{6}$$

$$y^{k+1} = y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \tag{7}$$

其中τ 为步长，通常取值于$\left(0, \frac{1+\sqrt{5}}{2}\right)$.

▶ ADMM的收敛准则: KKT条件

$$0 \in \partial_{x_1} L(x_1^*, x_2^*, y^*) = \partial f_1(x_1^*) + A_1^{\mathrm{T}} y^*, \tag{8a}$$

$$0 \in \partial_{x_2} L(x_1^*, x_2^*, y^*) = \partial f_2(x_2^*) + A_2^{\mathrm{T}} y^*, \tag{8b}$$

$$A_1 x_1^* + A_2 x_2^* = b, \tag{8c}$$

其中$L(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^{\mathrm{T}}(A_1 x_1 + A_2 x_2 - b)$. 条件(8c)又称为原始可行性条件，条件(8a)和条件(8b)又称为对偶可行性条件.

ADMM单步迭代最优性条件

- 由x_2的更新

$$x_2^k = \arg\min_x \left\{ f_2(x) + \frac{\rho}{2} \left\| A_1 x_1^k + A_2 x - b + \frac{y^{k-1}}{\rho} \right\|^2 \right\},$$

$$\Longrightarrow \qquad 0 \in \partial f_2(x_2^k) + A_2^{\mathrm{T}}[y^{k-1} + \rho(A_1 x_1^k + A_2 x_2^k - b)]. \tag{9}$$

当$\tau = 1$时, 由(7)可知 $0 \in \partial f_2(x_2^k) + A_2^{\mathrm{T}} y^k.$

- 由x_1的更新

$$x_1^k = \arg\min_x \left\{ f_1(x) + \frac{\rho}{2} \| A_1 x + A_2 x_2^{k-1} - b + \frac{y^{k-1}}{\rho} \|^2 \right\},$$

$$\Longrightarrow \qquad 0 \in \partial f_1(x_1^k) + A_1^{\mathrm{T}}[\rho(A_1 x_1^k + A_2 x_2^{k-1} - b) + y^{k-1}].$$

当$\tau = 1$时, 由(7)可知

$$0 \in \partial f_1(x_1^k) + A_1^{\mathrm{T}}(y^k + A_2(x_2^{k-1} - x_2^k)). \tag{10}$$

对比(10)和条件(8a)可知, 多出来的项为$A_1^{\mathrm{T}}A_2(x_2^{k-1}-x_2^k)$. 因此要检测对偶可行性只需要检测残差:

$$s^k = A_1^{\mathrm{T}}A_2(x_2^{k-1}-x_2^k).$$

- 当x_2更新取到精确解且$\tau=1$时, 判断ADMM 是否收敛只需要检测前述两个残差r^k, s^k是否充分小:

$$\begin{aligned} 0 &\approx \|r^k\| = \|A_1x_1^k + A_2x_2^k - b\| \quad \text{原始可行性}, \\ 0 &\approx \|s^k\| = \|A_1^{\mathrm{T}}A_2(x_2^{k-1}-x_2^k)\| \quad \text{对偶可行性}. \end{aligned} \tag{11}$$

▶ 典型优化问题(1)的ADMM的收敛准则:

$$\begin{aligned} 0 &\approx \|r^k\| = \|A_1x_1^k + A_2x_2^k - b\| \quad \textbf{原始可行} \\ 0 &\approx \|s^k\| = \|A_1^{\mathrm{T}}A_2(x_2^{k-1}-x_2^k)\| \quad \textbf{对偶可行} \end{aligned}$$

ADMM中的常用技巧: 线性化

▶ **对子问题目标函数进行二次近似, 使得子问题有显式解.**

- $$\min_{x_1} \quad f_1(x_1) + \frac{\rho}{2}\|A_1x_1 - v^k\|^2, \quad v^k = b - A_2x_2^k - \frac{1}{\rho}y^k. \tag{12}$$

- 当f可微时, 线性化将问题(12)变为

$$\begin{aligned} x_1^{k+1} &= \arg\min_{x_1}\left\{\left(\nabla f_1(x_1^k) + \rho A_1^{\mathrm{T}}(A_1x_1^k - v^k)\right)^{\mathrm{T}} x_1 + \frac{1}{2\eta_k}\|x_1 - x_1^k\|_2^2\right\} \\ &= x_1^k - \eta_k\left(\nabla f_1(x_1^k) + \rho A_1^T(A_1x_1^k - v^k)\right) \end{aligned}$$

其中η_k是步长参数. 【梯度下降步】

- 当f不可微但易于计算proximal算子时, 可以考虑只将二次项线性化:

$$\begin{aligned} x_1^{k+1} &= \arg\min_{x_1}\left\{f_1(x_1) + \rho\left(A_1^{\mathrm{T}}(A_1x_1^k - v^k)\right)^{\mathrm{T}} x_1 + \frac{1}{2\eta_k}\|x_1 - x_1^k\|_2^2\right\}. \\ &= \mathrm{prox}_{\eta_k f_1}\left(x_1^k - \eta_k\rho\left(A_1^{\mathrm{T}}(A_1x_1^k - v^k)\right)\right) \end{aligned}$$

【近端梯度步】

ADMM中的常用技巧: 缓存分解

- 若$f_1(x_1) = \frac{1}{2}\|Cx_1 - d\|_2^2$, 则$x_1$的更新(5)等价于求解线性方程组

$$(C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1)x_1 = C^{\mathrm{T}}d + \rho A_1^{\mathrm{T}} v^k.$$

- 首先对$C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1$进行Cholesky 分解并缓存分解的结果, 在每步迭代中只需要求解简单的三角形方程组.
- 当ρ发生更新时, 就要重新进行分解. 当$C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1$具有特殊结构(一部分容易求逆,另一部分低秩)时, 用Sherman-Morrison-Woodbury公式 求逆: $(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1}U)^{-1}V^T A^{-1}$

Sherman-Morrison-Woodbury 公式推导

证明：

$$M := \begin{pmatrix} I & -V^T \\ U & A \end{pmatrix} = \begin{pmatrix} I & 0 \\ U & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & A + UV^T \end{pmatrix} \begin{pmatrix} I & -V^T \\ 0 & I \end{pmatrix}$$

$$M = \begin{pmatrix} I & -V^T \\ U & A \end{pmatrix} = \begin{pmatrix} I & -V^T A^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} I + V^T A^{-1} U & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} I & 0 \\ A^{-1}U & I \end{pmatrix}$$

分别求M 的逆矩阵, 根据右下角分块相等得到SMW公式.

[1] William W. Hager, Updating the inverse of a matrix, SIAM Review, 31(2):221-239,1989

ADMM中的常用技巧: 优化转移

- 优化转移就是为了方便求解子问题, 可以用一个性质好的矩阵D近似二次项$A_1^{\mathrm{T}}A_1$, 此时子问题(12)替换为

$$\begin{aligned} x_1^{k+1} \quad = \quad & \arg\min_{x_1}\Big\{f_1(x_1)+\frac{\rho}{2}\|A_1x_1-v^k\|_2^2 \\ & +\frac{\rho}{2}(x_1-x_1^k)^{\mathrm{T}}(D-A_1^{\mathrm{T}}A_1)(x_1-x_1^k)\Big\}. \end{aligned}$$

- 通过选取合适的D, 当计算

$$\arg\min_{x_1}\Big\{f_1(x_1)+\frac{\rho}{2}x_1^{\mathrm{T}}Dx_1\Big\}$$

明显比计算$\arg\min\limits_{x_1}\{f_1(x_1)+\frac{\rho}{2}x_1^{\mathrm{T}}A_1^{\mathrm{T}}A_1x_1\}$ 要容易时, 优化转移可以极大地简化子问题的计算. (1) 当$D=\frac{\eta_k}{\rho}I$时, 优化转移等价于做单步的近似点梯度步; (2) $D=\sum_{i=1}^{r}(\lambda_i-\lambda_r)u_iu_i^T+\lambda_rI$, 其中$A_1^TA_1=\sum_{i=1}^{m}\lambda_iu_iu_i^T$ 为谱分解.

ADMM中的常用技巧: 二次罚项系数的动态调节

▶ 动态调节二次罚项系数在交替方向乘子法的实际应用中是一个非常重要的数值技巧.

- 由(11)知, 求解过程中二次罚项系数ρ太大会导致原始可行性$\|r^k\|$下降很快, 但是对偶可行性$\|s^k\|$下降很慢; 二次罚项系数太小, 则效果相反. 这都会导致收敛比较慢或得到的解的可行性很差. 一个自然的想法是在每次迭代时动态调节惩罚系数ρ的大小, 从而使得原始可行性和对偶可行性能够以比较一致的速度下降到零.

- 简单有效的动态调节二次罚项系数的方式:

$$\rho^{k+1} = \begin{cases} \gamma_p \rho^k, & \|r^k\| > \mu\|s^k\|, \\ \rho^k/\gamma_d & \|s^k\| > \mu\|r^k\|, \\ \rho^k, & \text{otherwise}, \end{cases}$$

其中$\mu > 1, \gamma_p > 1, \gamma_d > 1$是参数, 常见的选择为$\mu = 10, \gamma_p = \gamma_d = 2$. 在迭代过程中将原始可行性$\|r^k\|$和对偶可行性$\|s^k\|$保持在彼此的$\mu$倍内.

ADMM中的常用技巧: 超松弛

考虑求解优化问题(1)的ADMM迭代格式:

$$x_1^{k+1} = \arg\min_{x_1} L_\rho(x_1, x_2^k, y^k), \tag{13}$$

$$x_2^{k+1} = \arg\min_{x_2} L_\rho(x_1^{k+1}, x_2, y^k), \tag{14}$$

$$y^{k+1} = y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b). \tag{15}$$

- 在(14)与(15)中, $A_1 x_1^{k+1}$可以被替换为

  $$\alpha_k A_1 x_1^{k+1} + (1-\alpha_k)(A_2 x_2^k - b),$$

  其中$\alpha_k \in (0,2)$是一个松弛参数.

- 当$\alpha_k > 1$时, 这种技巧称为超松弛. 当$\alpha_k < 1$时, 这种技巧称为欠松弛. 实验表明$\alpha_k \in [1.5, 1.8]$的超松弛可以提高收敛速度.

ADMM应用：LASSO与广义LASSO

LASSO 问题

$$\min \quad \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|^2 \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,z} & \frac{1}{2}\|Ax-b\|^2 + \mu\|z\|_1, \\ \text{s.t.} & x = z. \end{cases}$$

- 交替方向乘子法迭代格式为

$$\begin{aligned} x^{k+1} &= \arg\min_x \left\{ \frac{1}{2}\|Ax-b\|^2 + \frac{\rho}{2}\|x - z^k + y^k/\rho\|_2^2 \right\}, \\ &= (A^{\mathrm{T}}A + \rho I)^{-1}(A^{\mathrm{T}}b + \rho z^k - y^k), \\ z^{k+1} &= \arg\min_z \left\{ \mu\|z\|_1 + \frac{\rho}{2}\|x^{k+1} - z + y^k/\rho\|^2 \right\}, \\ &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(x^{k+1} + y^k/\rho\right), \\ y^{k+1} &= y^k + \tau\rho(x^{k+1} - z^{k+1}). \end{aligned}$$

- 因为$\rho > 0$, 所以$A^{\mathrm{T}}A + \rho I$总是可逆的．x迭代本质上是计算一个岭回归问题(ℓ_2范数平方正则化的最小二乘问题). 而对z的更新为ℓ_1范数的邻近算子, 同样有显式解．在求解x迭代时，若使用固定的罚因子ρ, 我们可以缓存矩阵$A^{\mathrm{T}}A + \rho I$的初始分解, 从而减小后续迭代中的计算量.
- 在LASSO 问题中, 矩阵$A \in \mathbb{R}^{m\times n}$通常有较多的列$m \ll n$, 因此$A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$是一个低秩矩阵，二次罚项的作用就是将$A^{\mathrm{T}}A$增加了一个正定项．该ADMM 主要运算量来自更新$x$变量时求解线性方程组，复杂度为$O(n^3)$. 若使用缓存分解技术或SMW 公式:

$$(A^{\mathrm{T}}A + \rho I_n)^{-1} = \frac{1}{\rho}\left[I_n - A^{\mathrm{T}}(AA^{\mathrm{T}} + \rho I_m)^{-1}A\right],$$

其中的求逆运算复杂度降为$O(m^3)$.

■ **考虑LASSO 问题的对偶问题**:

$$\begin{aligned} \min \quad & b^{\mathrm{T}} y + \tfrac{1}{2}\|y\|^2, \\ \text{s.t.} \quad & \|A^{\mathrm{T}} y\|_\infty \le \mu. \end{aligned} \tag{16}$$

- 引入约束$A^{\mathrm{T}}y + z = 0$, 则

$$(16) \iff \left\{ \begin{aligned} \min \quad & \underbrace{b^{\mathrm{T}} y + \frac{1}{2}\|y\|^2}_{f(y)} + \underbrace{I_{\|z\|_\infty \le \mu}(z)}_{h(z)}, \\ \text{s.t.} \quad & A^{\mathrm{T}} y + z = 0. \end{aligned} \right. \tag{17}$$

- 对偶问题(17)的增广拉格朗日函数为

$$L_\rho(y, z, x) = b^{\mathrm{T}} y + \frac{1}{2}\|y\|^2 + I_{\|z\|_\infty \le \mu}(z) - x^{\mathrm{T}}(A^{\mathrm{T}} y + z) + \frac{\rho}{2}\|A^{\mathrm{T}} y + z\|^2.$$

- 当固定y, x时, 对z的更新即向无穷范数球$\{z \mid \|z\|_\infty \le \mu\}$做欧几里得投影, 将每个分量截断在区间$[-\mu, \mu]$中. 当固定$z, x$时, 对$y$的更新即求解线性方程组

$$(I + \rho A A^{\mathrm{T}}) y = A(x^k - \rho z^{k+1}) - b. \tag{18}$$

- LASSO 问题的对偶问题(17)的ADMM 迭代格式为

$$
\begin{aligned}
z^{k+1} &= \mathcal{P}_{\|z\|_\infty \le \mu}\left(x^k/\rho - A^{\mathrm{T}}y^k\right), \\
y^{k+1} &= (I+\rho AA^{\mathrm{T}})^{-1}\Big(A(x^k-\rho z^{k+1})-b\Big), \\
x^{k+1} &= x^k - \tau\rho(A^{\mathrm{T}}y^{k+1}+z^{k+1}).
\end{aligned}
$$

  - 虽然ADMM 应用于对偶问题也需要求解一个线性方程组(18), 但由于LASSO 问题的特殊性$m \ll n$, 求解y更新的线性方程组(18)需要的计算量是$O(m^3)$, 使用缓存分解技巧后可进一步降低至$O(m^2)$, 这大大小于针对原始问题的ADMM.

- **广义LASSO 问题**指x 本身不稀疏, 但在某种变换下是稀疏的:

$$
\min_x \quad \mu\|Fx\|_1 + \frac{1}{2}\|Ax-b\|^2. \tag{19}
$$

- 当$F \in \mathbb{R}^{(n-1)\times n}$是一阶差分矩阵

$$F_{ij} = \begin{cases} 1, & j = i+1, \\ -1, & j = i, \\ 0, & \text{otherwise}, \end{cases}$$

且$A = I$时, 广义LASSO问题(19)为图像去噪问题的TV 模型:

$$\min_{x} \quad \frac{1}{2}\|x-b\|^2 + \mu \sum_{i=1}^{n-1} |x_{i+1} - x_i|.$$

当$A = I$且F是二阶差分矩阵时, 问题(19)被称为一范数趋势滤波.

▶ 广义LASSO问题(19)等价于

$$\begin{aligned} \min_{x,z} \quad & \frac{1}{2}\|Ax-b\|^2 + \mu\|z\|_1 \\ \text{s.t.} \quad & Fx - z = 0, \end{aligned} \tag{20}$$

- 对广义LASSO问题(20)的ADMM迭代为

$$
\begin{aligned}
x^{k+1} &= (A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F)^{-1}\left(A^{\mathrm{T}}b + \rho F^{\mathrm{T}}\left(z^k - \frac{y^k}{\rho}\right)\right),\\
z^{k+1} &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(Fx^{k+1} + \frac{y^k}{\rho}\right),\\
y^{k+1} &= y^k + \tau\rho(Fx^{k+1} - z^{k+1}).
\end{aligned}
$$

- 注: 对于全变差去噪问题, $A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F$是三对角矩阵, 此时$x$迭代可以在$\mathcal{O}(n)$的时间复杂度内解决；对于图像去模糊问题，$A$ 是卷积算子，则利用傅里叶变换可将求解方程组的复杂度降低至$O(n\log n)$；对于一范数趋势滤波问题，$A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F$是五对角矩阵，所以$x$迭代仍可以在$\mathcal{O}(n)$的时间复杂度内解决.

ADMM应用: 稀疏逆协方差矩阵估计

■ 稀疏逆协方差矩阵估计问题:

$$\min_{X} \quad \langle S, X\rangle - \ln\det X + \mu\|X\|_1, \tag{21}$$

其中S是已知的对称矩阵, 通常由样本协方差矩阵得到. 变量$X \in \mathcal{S}^n_{++}$, $\|\cdot\|_1$定义为矩阵所有元素绝对值的和.

- (21)的目标函数由光滑项和非光滑项组成, 将问题的两部分分离:

$$\begin{aligned} \min \quad & \underbrace{\langle S, X\rangle - \ln\det X}_{f(X)} + \underbrace{\mu\|Z\|_1}_{h(Z)}, \\ \text{s.t.} \quad & X = Z. \end{aligned}$$

- 其增广拉格朗日函数为

$$L_\rho(X, Z, U) = \langle S, X\rangle - \ln\det X + \mu\|Z\|_1 + \langle U, X - Z\rangle + \frac{\rho}{2}\|X - Z\|_F^2.$$

- ADMM迭代:
  - 对X的更新, 固定Z^k, U^k, 关于X的子问题是凸的, 故由最优性条件

$$S - X^{-1} + U^k + \rho(X - Z^k) = 0,$$

$$\Longrightarrow \quad X^{k+1} = Q\mathrm{Diag}(x_1, x_2, \cdots, x_n)Q^{\mathrm{T}},$$

其中Q包含矩阵$S - \rho Z^k + U^k$的所有特征向量, x_i的表达式为

$$x_i = \frac{-d_i + \sqrt{d_i^2 + 4\rho}}{2\rho},$$

d_i为矩阵$S - \rho Z^k + U^k$的第i个特征值.

  - 对Z的更新, 固定X^{k+1}, U^k, 关于Z的更新为矩阵ℓ_1 范数的邻近算子.
  - 乘子更新: $U^{k+1} = U^k + \tau\rho(X^{k+1} - Z^{k+1})$.

ADMM应用: 矩阵分离问题

■ **矩阵分离问题:**

$$\begin{aligned} \min_{X,S} \quad & \|X\|_* + \mu\|S\|_1 \\ \textbf{s.t.} \quad & X + S = M, \end{aligned} \tag{22}$$

其中$\|\cdot\|_1$与$\|\cdot\|_*$分别表示矩阵ℓ_1范数与核范数.

- 问题(22)的增广拉格朗日函数

$$L_\rho(X, S, Y) = \|X\|_* + \mu\|S\|_1 + \langle Y, X + S - M\rangle + \frac{\rho}{2}\|X + S - M\|_F^2.$$

- 对X的更新

$$
\begin{aligned}
X^{k+1} &= \arg\min_{X}\ L_\rho(X, S^k, Y^k) \\
&= \arg\min_{X} \left\{ \|X\|_* + \frac{\rho}{2} \left\| X + S^k - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\
&= \arg\min_{X} \left\{ \frac{1}{\rho}\|X\|_* + \frac{1}{2} \left\| X + S^k - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\
&= U\mathrm{Diag}\Big(\mathrm{prox}_{(1/\rho)\|\cdot\|_1}(\sigma(A))\Big)V^{\mathrm{T}},
\end{aligned}
$$

其中$A = M - S^k - \frac{Y^k}{\rho}$, $\sigma(A)$为A的所有非零奇异值构成的向量并且$U\mathrm{Diag}(\sigma(A))V^{\mathrm{T}}$为$A$的奇异值分解.

- 对S的更新

$$
\begin{aligned}
S^{k+1} &= \arg\min_{S}\ L_\rho(X^{k+1}, S, Y^k) \\
&= \arg\min_{S} \left\{ \mu\|S\|_1 + \frac{\rho}{2}\left\| X^{k+1} + S - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\
&= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(M - X^{k+1} - \frac{Y^k}{\rho} \right).
\end{aligned}
$$

▶ 矩阵分离问题(22)的ADMM迭代:

$$
\begin{aligned}
X^{k+1} &= U\mathrm{Diag}\Big(\mathrm{prox}_{(1/\rho)\|\cdot\|_1}(\sigma(A))\Big)V^{\mathrm{T}}, \\
S^{k+1} &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(M - L^{k+1} - \frac{Y^k}{\rho} \right), \\
Y^{k+1} &= Y^k + \tau\rho(X^{k+1} + S^{k+1} - M).
\end{aligned}
$$

ADMM应用: 全局一致性优化问题

全局一致性优化问题:

$$\min_{x_i, z} \sum_{i=1}^{N} \phi_i(x_i) \quad \text{s.t.} \quad x_i - z = 0,\ i = 1, 2, \cdots, N.$$

- 增广拉格朗日函数

$$L_\rho(x_1, \cdots, x_N, z, y_1, \cdots, y_N) = \sum_{i=1}^{N} \phi_i(x_i) + \sum_{i=1}^{N} y_i^{\mathrm{T}}(x_i - z) + \frac{\rho}{2} \sum_{i=1}^{N} \|x_i - z\|^2.$$

注: 虽然表面上看增广拉格朗日函数有$(N+1)$个变量块, 但本质上还是两个变量块. 这是因为在更新某个x_i时并没有利用其他x_i, 所有x_i可以看成一个整体. 相应地, 所有乘子y_i也可以看成一个整体.

- 全局一致性优化问题的ADMM迭代:

$$
\begin{aligned}
x_i^{k+1} &= \mathrm{prox}_{\phi_i/\rho}\left(z^k - y_i^k/\rho\right), \ i = 1, 2, \cdots, N, \\
z^{k+1} &= \frac{1}{N}\sum_{i=1}^{N}\left(x_i^{k+1} + y_i^k/\rho\right), \\
y_i^{k+1} &= {y_i}^k + \tau\rho({x_i}^{k+1} - z^{k+1}), \ i = 1, 2, \cdots, N.
\end{aligned}
$$