

最优化方法 第七讲

Weiwen Wang(王伟文)

暨南大学

2025 年 5 月 6 日

目录

- 1 带约束优化问题的最优性条件
- 2 Krein-Milman 定理
- 3 Frank-Wolfe 算法
 - Frank-Wolfe 算法特性
 - 收敛性分析
 - 矩阵补全
 - Lasso
 - 仿射不变收敛性

本讲考虑带约束优化问题 (P)

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } \mathbf{x} \in C \end{aligned}$$

其中 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 为连续可微函数, C 为有界闭集.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

定义 1 (极端点 Extreme Points)

设 $S \subset \mathbb{R}^d$. $\mathbf{z} \in S$ 称为 S 的极端点, 若不存在 $\mathbf{x}, \mathbf{y} \in S (\mathbf{x} \neq \mathbf{y})$ 及 $\lambda \in (0, 1)$ 使得 $\mathbf{z} = \lambda \mathbf{x} + (1 - \lambda) \mathbf{y}$. 集合 S 的所有极端点记为 $\text{ext}(S)$.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

定理 2 (Krein-Milman 定理)

若 $S \subset \mathbb{R}^d$ 为紧凸集, 则

$$S = \text{conv}(\text{ext}(S)).$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

命题 3

设 $f: C \subset \mathbb{R}^d \rightarrow \mathbb{R}$ 为紧凸集上的连续凸函数, 则 f 在 C 至少存在一个最大值点且为极 endpoint.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

1. 由 Weierstrass 定理, f 在 C 上的最大值点 $\mathbf{x}^*$ 存在.
2. 由 Krein-Milman 定理知, 存在 $k \in \mathbb{Z}_+$, $\lambda \in \Delta_k$ 使得

$$\mathbf{x}^* = \sum_{i=1}^k \lambda_i \mathbf{z}_i, \quad \exists \{\mathbf{z}_i\}_{i \in [d]} \subset \text{ext}(C).$$

由凸函数性质及 $\mathbf{x}^*$ 的最优性

$$f(\mathbf{x}^*) = f\left(\sum_{i=1}^k \lambda_i \mathbf{z}_i\right) \leq \sum_{i=1}^k \lambda_i f(\mathbf{z}_i) \leq \max_{\mathbf{z}_i: i \in [d]} f(\mathbf{z}_i) = f(\mathbf{x}^*).$$

故存在 $j \in [d]$, 使得 $\mathbf{z}_j \in \{\mathbf{z}_i\}_{i \in [d]} \subset \text{ext}(C)$ 为 f 在 C 上的最大值点. □

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

设定初始点 $\mathbf{x}_0$ 、迭代步长 $\gamma_k \in (0, 1)$.

- $\mathbf{z}_k = \arg \min_{\mathbf{x} \in \mathbf{C}} \{f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x} \rangle\} = \arg \min_{\mathbf{x} \in \mathbf{C}} \langle \nabla f(\mathbf{x}_k), \mathbf{x} \rangle$
- $\mathbf{x}_{k+1} = (1 - \gamma_k)\mathbf{x}_k + \mathbf{z}_k$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

Frank-Wolfe 算法特性

- 若 C 为凸集, $\mathbf{x}_0 \in C$, 则 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \in C$.
- 与投影梯度法比较, 每一步只需要求解带约束的线性规划问题.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

引理 4

设 $\mathbf{x}^*$ 是优化问题 P 的最优解, f 为 C 上的凸函数, 则

$$\langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle \geq f(\mathbf{x}_k) - f(\mathbf{x}^*)$$

- 这表明 $\langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle$ 可作为最优性的上界估计条件.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

由 f 的凸性

$$\langle f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}^* \rangle \geq f(\mathbf{x}_k) - f(\mathbf{x}^*)$$

又因为 $\mathbf{z}_k = \operatorname{argmin}_{\mathbf{x} \in C} \langle \nabla f(\mathbf{x}_k), \mathbf{x} \rangle$, 故

$$\langle \nabla f(\mathbf{x}_k), \mathbf{z}_k \rangle \leq \langle \nabla f(\mathbf{x}_k), \mathbf{x}^* \rangle$$

综上可得

$$\langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle \geq \langle f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}^* \rangle \geq f(\mathbf{x}_k) - f(\mathbf{x}^*)$$



带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

引理 5

设 f 是 L -smooth, C 为凸集, 则 Frank-Wolfe 算法的迭代过程满足

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \gamma_k \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle + \frac{L\gamma_k^2}{2} \|\mathbf{z}_k - \mathbf{x}_k\|^2 \quad k \geq 0.$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

L -smooth 函数性质

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2$$

代入 $\mathbf{x}_{k+1} - \mathbf{x}_k = \gamma_k(\mathbf{z}_k - \mathbf{x}_k)$ 即完成证明.



带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

定理 6

设 f 是凸函数且 L -smooth, C 为紧凸集, 记 $\mathbf{x}^*$ 为优化问题 (P) 的最优解, 取 $\gamma_k = 2/(k+2)$, 则 Frank-Wolfe 算法满足

$$f(\mathbf{x}_K) - f(\mathbf{x}^*) \leq \frac{2LD^2}{K+1} \quad K \geq 1.$$

其中 $D = \max_{\mathbf{u}, \mathbf{v} \in C} \|\mathbf{u} - \mathbf{v}\|$.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

由引理 5 及引理 4

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) - \gamma_k \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle + \frac{L\gamma_k^2}{2} \|\mathbf{z}_k - \mathbf{x}_k\|^2 \\ &\leq f(\mathbf{x}_k) - \gamma_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) + \frac{L\gamma_k^2 D^2}{2} \end{aligned}$$

整理得到

$$f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*) \leq (1 - \gamma_k)(f(\mathbf{x}_k) - f(\mathbf{x}^*)) + \frac{L\gamma_k^2 D^2}{2}$$

记 $\Delta_k = f(\mathbf{x}_k) - f(\mathbf{x}^*)$

□

证明.

即有递推关系

$$\begin{aligned}\Delta_{k+1} &\leq (1 - \gamma_k)\Delta_k + \frac{L\gamma_k^2 D^2}{2} \\ &= \frac{k}{k+2}\Delta_k + \frac{2LD^2}{(k+2)^2}\end{aligned}$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

即有递推关系

$$\begin{aligned}\Delta_{k+1} &\leq (1 - \gamma_k)\Delta_k + \frac{L\gamma_k^2 D^2}{2} \\ &= \frac{k}{k+2}\Delta_k + \frac{2LD^2}{(k+2)^2}\end{aligned}$$

当 $k = 0$ 时, $\Delta_1 \leq \frac{2LD^2}{4} \leq \frac{2LD^2}{2}$, 设有

$$\Delta_k \leq \frac{2LD^2}{k+1} \quad k \geq 1.$$

$$\begin{aligned}\Delta_{k+1} &\leq \frac{k}{k+2} \frac{2LD^2}{k+1} + \frac{2LD^2}{(k+2)^2} \\ &\leq \frac{2k \cdot LD^2}{(k+1)(k+2)} + \frac{2LD^2}{(k+2)(k+1)} \leq \frac{2LD^2}{k+2}.\end{aligned}$$

□

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

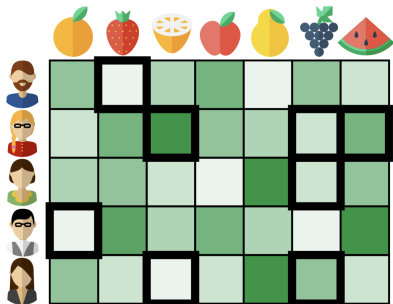
收敛性分析

矩阵补全

Lasso

仿射不变收敛性

矩阵补全



带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

$\Omega = \{Person\} \times \{Item\}$ 表示具有观测打分匹配关系. Y 表示观测打分矩阵, 其中 $Y_{\overline{\Omega}} = [0]_{\overline{\Omega}}$. 矩阵补全通过恢复矩阵 $X \in \mathbb{R}^{p \times t}$ 获得所有配对关系的打分, 即求解以下优化问题

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

$\Omega = \{Person\} \times \{Item\}$ 表示具有观测打分匹配关系. Y 表示观测打分矩阵, 其中 $Y_{\Omega} = [0]_{\Omega}$. 矩阵补全通过恢复矩阵 $X \in \mathbb{R}^{p \times t}$ 获得所有配对关系的打分, 即求解以下优化问题

$$\min_{X \in \mathbb{R}^{p \times t}} \sum_{i,j \in \Omega} (X_{ij} - Y_{ij})^2 \quad \text{s.t. } \text{rank}(X) \leq r.$$

可松弛为

$$\begin{aligned} \min_{X \in \mathbb{R}^{p \times t}} \quad & \frac{1}{2} \|\Pi_{\Omega}(X) - Y\|_F^2 \\ \text{s.t.} \quad & \|X\|_* \leq r. \end{aligned}$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

$$f(X) \triangleq \frac{1}{2} \|\Pi_{\Omega}(X) - Y\|_F^2$$

$$\nabla f(X) = \Pi_{\Omega}(X) - Y$$

应用 Frank-Wolfe 算法求解需要计算

$$\min_Z \langle \nabla f(X), Z \rangle \triangleq \text{tr}(\nabla f(X)^T Z)$$

$$\text{s.t. } \|Z\|_* \leq r$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

命题 7

设 $A \in \mathbb{R}^{d_1 \times d_2}$, 其奇异值分解 (*Singular Value Decomposition, SVD*) 为 $A = U\Sigma V^T$. 则优化问题

$$\min \langle A, X \rangle \quad \text{s.t.} \quad \|X\|_* \leq r$$

的最优解为 $X^* = -r\mathbf{u}_1\mathbf{v}_1^T$, 其中 $\mathbf{u}_1$ 和 $\mathbf{v}_1$ 分别为 U 、 V 的第一列.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

证明.

由矩阵 Hölder 不等式

$$|\langle A, X \rangle| \leq \|A\|_2 \cdot \|X\|_*$$

故

$$\langle A, X \rangle \geq -\|A\|_2 \cdot \|X\|_* \geq r\|A\|_2$$

当 $X = -r\mathbf{u}_1\mathbf{v}_1^T$ 时, $\|X\|_* = r$ 且上述不等式的等号成立.

□

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

```
import numpy as np
from scipy.sparse.linalg import svds

def matrix_completion(X0, Omega, Y, r, iteration):
    for i in range(iteration):
        gradient = projection(X, Omega) - Y
        u, sigma, vT = svds(gradient, k=1)
        V = -r * u @ vT
        X = i/(i+2) * X + 2/(k+2) * V
    return X
```


Lasso

$$\begin{aligned} \min_{\boldsymbol{\theta} \in \mathbb{R}^d} f(\boldsymbol{\theta}) &\triangleq \frac{1}{2} \|\mathbf{A}\boldsymbol{\theta} - \mathbf{y}\|^2 \\ \text{s.t. } \|\boldsymbol{\theta}\|_1 &\leq \tau. \end{aligned}$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

命题 8

设 $\mathbf{u} \in \mathbb{R}^d$, 优化问题

$$\min_{\mathbf{v} \in \mathbb{R}^d} \langle \mathbf{u}, \mathbf{v} \rangle \quad \text{s.t. } \|\mathbf{v}\|_1 \leq r.$$

的最优解为 $\mathbf{v}^* = -r \operatorname{sgn}(u_{k^*}) \mathbf{e}_{k^*}$, 其中 $k^* = \operatorname{argmax}_{k \in [d]} |u_k|$.

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

```
import numpy as np
def lasso_fw_sovler(A, b, x, r, max_iter):
    for t in range(max_iter):
        g = A.T @ (A @ x - b)
        idx = np.argsort(-np.abs(g))[0]
        s = np.zeros_like(x)
        s[idx] = - r * np.sign(g[idx])
        x = t/(t+2) * x + 2/(t+2) * s
    return x
```

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

步长策略

- 精确线搜索

$$\gamma_k = \operatorname{argmin}_{\gamma \in [0,1]} f((1-\gamma)\mathbf{x}_k + \gamma\mathbf{z}_k)$$

- 上界最小化 由 L -smooth 性质

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \gamma \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle + \frac{L\gamma^2}{2} \|\mathbf{z}_k - \mathbf{x}_k\|^2$$

故

$$\gamma_k = \operatorname{argmin}_{\gamma \in [0,1]} \left\{ f(\mathbf{x}_k) - \gamma \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle + \frac{L\gamma^2}{2} \|\mathbf{z}_k - \mathbf{x}_k\|^2 \right\}$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

$$\gamma_k = \arg \min_{\gamma \in [0,1]} \left\{ f(\mathbf{x}_k) - \gamma \langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle + \frac{L\gamma^2}{2} \|\mathbf{z}_k - \mathbf{x}_k\|^2 \right\}$$

注意到 $f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle \geq f(\mathbf{x}_k) - f(\mathbf{x}^*) \geq 0$, 故解得

$$\gamma_k = \min \left(1, \frac{\langle f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle}{L \|\mathbf{z}_k - \mathbf{x}_k\|^2} \right)$$

仿射不变收敛性

定义

$$\mathcal{G}_{(f,C)} = \sup_{\substack{\mathbf{x}, \mathbf{z} \in C, \gamma \in (0, 1] \\ \mathbf{y} = (1 - \gamma)\mathbf{x} + \gamma\mathbf{z}}} \frac{1}{\gamma^2} (f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle)$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

引理 9

设 f 在集合 C 上为可微凸函数且为 L -smooth, 则

$$\mathcal{G}_{(f,C)} \leq \frac{L}{2} D^2$$

其中 $D = \max_{\mathbf{u}, \mathbf{v} \in C} \|\mathbf{u} - \mathbf{v}\|$.

证明.

对任意 $\mathbf{x}, \mathbf{z} \in \mathcal{C}$, $\gamma \in (0, 1]$, 定义 $\mathbf{y} = (1 - \gamma)\mathbf{x} + \gamma\mathbf{z}$.

由 L -smooth 性质

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle &\leq \frac{L}{2} \|\mathbf{x} - \mathbf{y}\|^2 \\ &= \frac{L\gamma^2}{2} \|\mathbf{x} - \mathbf{z}\|^2 \\ &\leq \frac{L\gamma^2}{2} D^2 \end{aligned}$$

整理得到

$$\frac{1}{\gamma^2} (f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle) \leq \frac{L}{2} D^2$$

不等式两端同时关于 $\mathbf{x}, \mathbf{z}$ 和 γ 取最大值, 即有

$$\mathcal{G}_{(f, \mathcal{C})} \leq \frac{L}{2} D^2$$

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性

定理 10

设 f 是可微凸函数, C 为紧凸集, 记 $\mathbf{x}^*$ 为优化问题 (P) 的最优解, 取 $\gamma_k = 2/(k+2)$, 则 Frank-Wolfe 算法满足

$$f(\mathbf{x}_K) - f(\mathbf{x}^*) \leq \frac{4\mathcal{G}(f,C)}{K+1} \quad K \geq 1.$$

证明.

由 $\mathcal{G}_{(f,C)}$ 定义知

$$\frac{1}{\gamma_k^2} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) - \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle) \leq \mathcal{G}_{(f,C)}$$

整理得到

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \gamma_k^2 \mathcal{G}_{(f,C)} \\ &= f(\mathbf{x}_k) + \gamma_k \langle \nabla f(\mathbf{x}_k), \mathbf{z}_k - \mathbf{x}_k \rangle + \gamma_k^2 \mathcal{G}_{(f,C)} \\ &\leq f(\mathbf{x}_k) + \gamma_k (f(\mathbf{x}^*) - f(\mathbf{x}_k)) + \gamma_k^2 \mathcal{G}_{(f,C)} \end{aligned}$$

其中最后一个不等号来自不等式 $\langle \nabla f(\mathbf{x}_k), \mathbf{x}_k - \mathbf{z}_k \rangle \geq f(\mathbf{x}_k) - f(\mathbf{x}^*)$.

定义 $\Delta_k = f(\mathbf{x}_k) - f(\mathbf{x}^*)$, 有

$$\Delta_{k+1} \leq (1 - \gamma_k) \Delta_k + \gamma_k^2 \mathcal{G}_{(f,C)}.$$

全下部分证明同定理 6

□

带约束优化问题的最优性条件

Krein-Milman 定理

Frank-Wolfe 算法

Frank-Wolfe 算法特性

收敛性分析

矩阵补全

Lasso

仿射不变收敛性