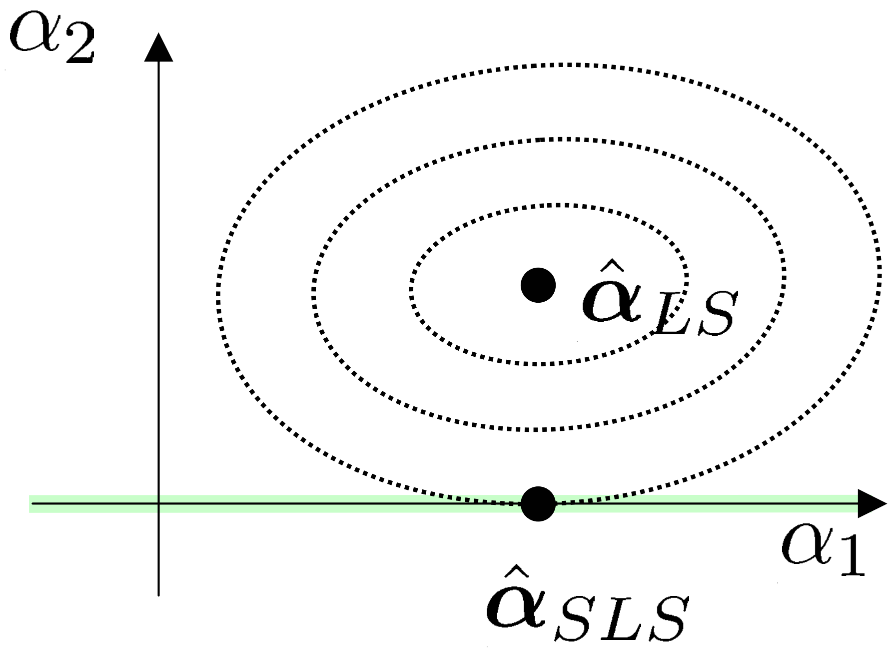
Sparseness of Solution

- In SLS, if the subspace is spanned by a subset of basis functions $\{\varphi_i(\boldsymbol{x})\}_{i=1}^{p}$, some of the parameters $\{\alpha_i\}_{i=1}^{p}$ are zero.



Model Choice

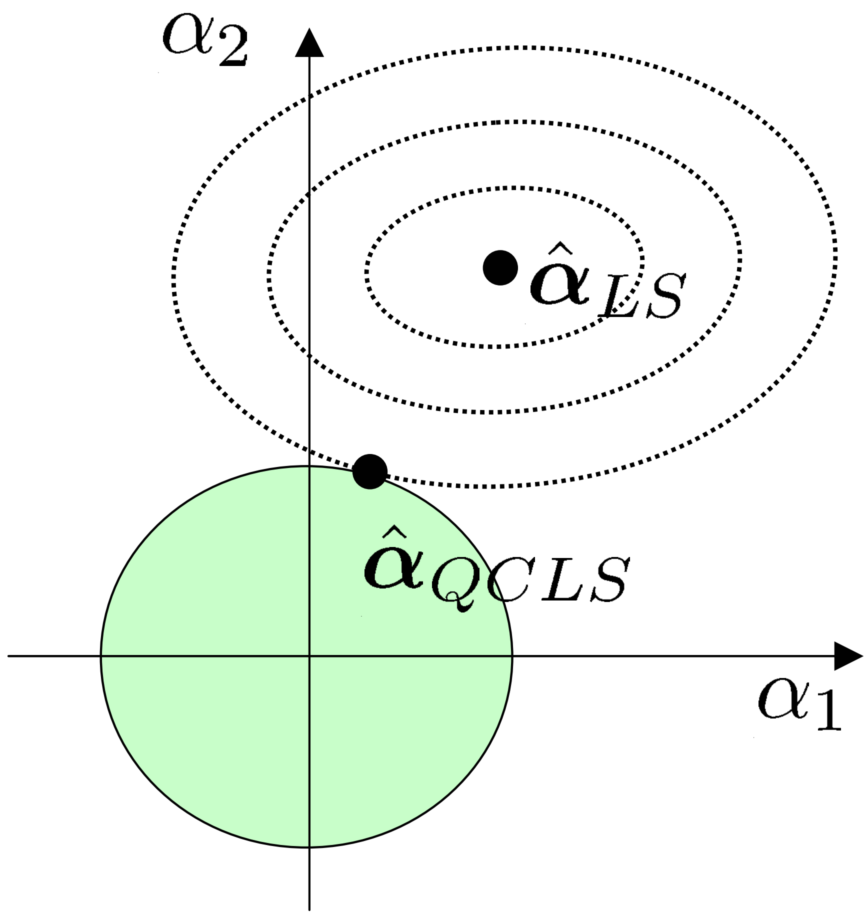
- Sparse solution is computationally advantageous in calculating the output values.

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{p} \alpha_i \varphi_i(\boldsymbol{x})$$

- However, the choice of such subspaces is discrete.
- Combinatorial explosion! 2^p

Property of QCLS

- In QCLS, model choice is continuous: λ
- However, solution is not generally sparse.



Learning Methods

- Sparse learning
- How to solve optimization problem

Non-Linear Learning for Linear and Kernel Models

- Linear/kernel models

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{p} \alpha_i \varphi_i(\boldsymbol{x}) \qquad \hat{f}(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i)$$

- Non-linear learning

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{L}(\boldsymbol{y})$$

$\boldsymbol{L}$:Non-linear function

Sparseness and Continuous Model Choice

- Two approaches to avoiding over-fitting:
  - Subspace LS

    Sparse but discrete model choice
  - Quadratically constrained LS

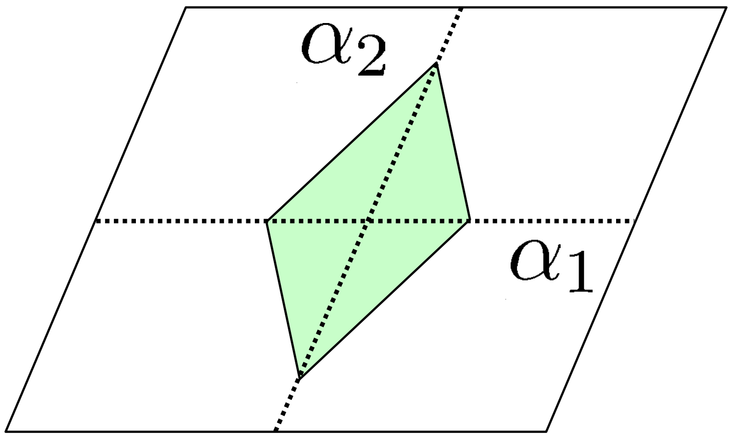
    Continuous model choice but non-sparse
- We want to have sparseness and continuous model choice at the same time.

l1-Constrained LS

- Restrict the search space within a hyper-rhombus.

$$\hat{\boldsymbol{\alpha}}_{\ell_1 CLS} = \operatorname*{argmin}_{\boldsymbol{\alpha}} J_{LS}(\boldsymbol{\alpha})$$

$$\text{subject to } \|\boldsymbol{\alpha}\|_1 \leq C$$

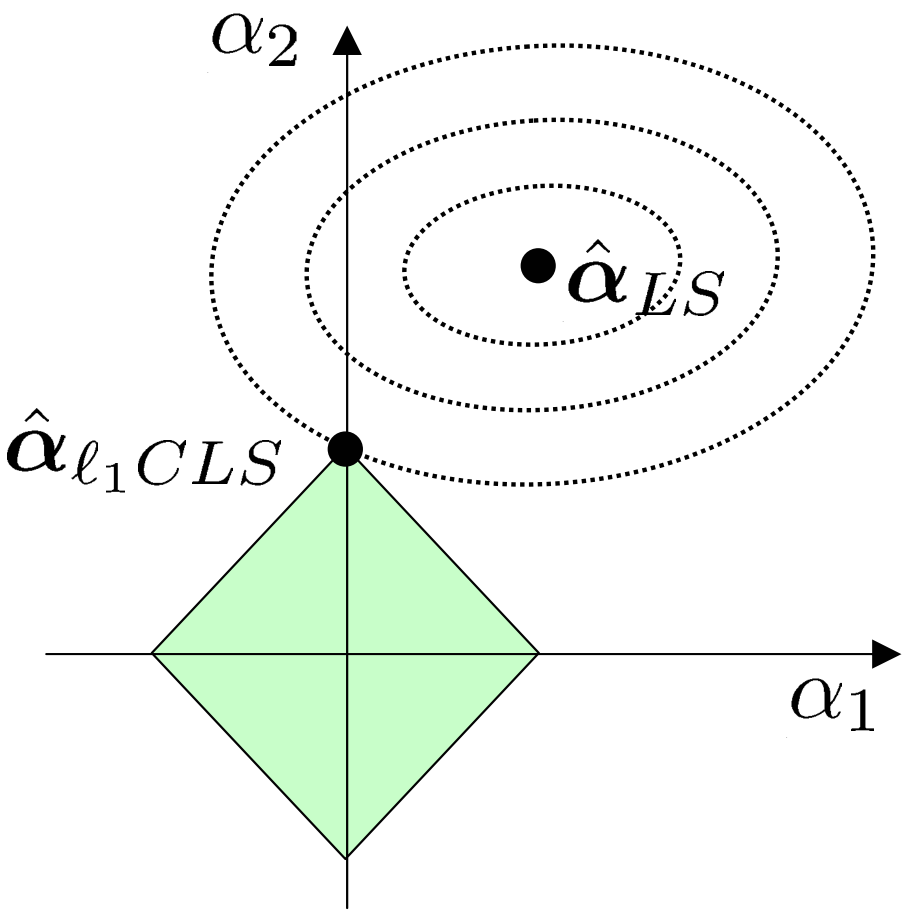


$\ell_1 - \text{norm}$

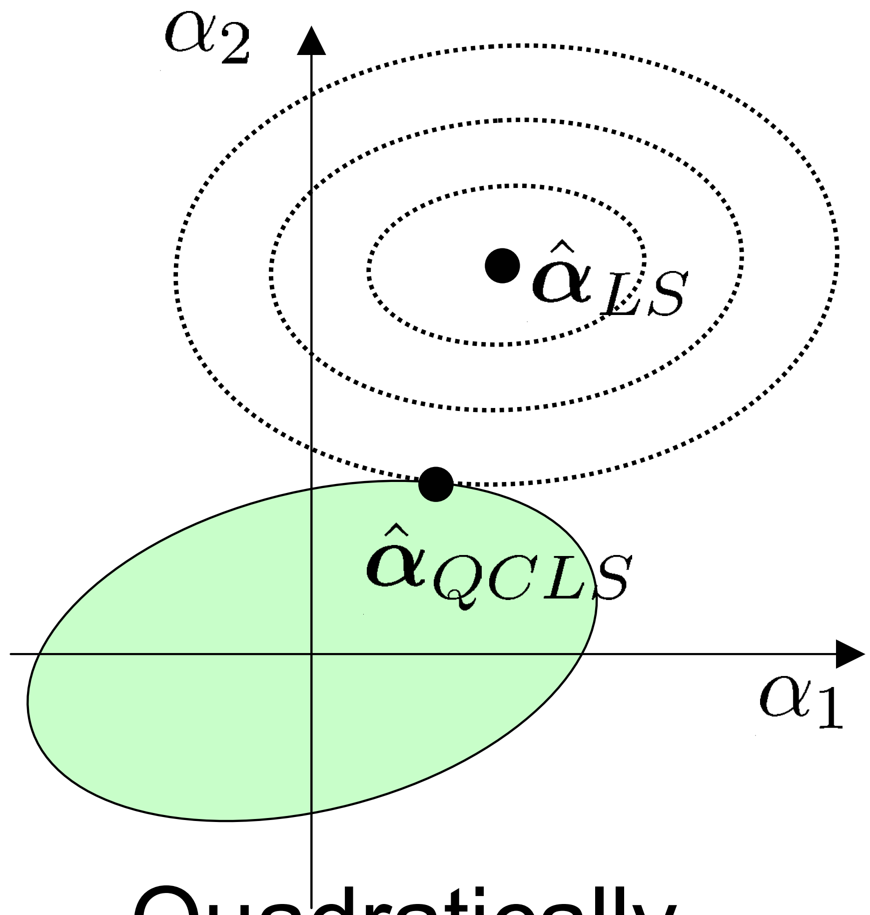
$$\|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^{p} |\alpha_i|$$

Why Sparse?

- The solution often touches on an axis.



ℓ_1 constrained LS



Quadratically constrained LS

How to Obtain Solutions

- Lagrangian:

$$J_{\ell_1 CLS}(\boldsymbol{\alpha}) = J_{LS}(\boldsymbol{\alpha}) + \lambda \|\boldsymbol{\alpha}\|_1$$

- $\lambda\ (\geq 0)$:Lagrange multiplier
- In practice, we start from $\lambda\ (\geq 0)$ and solve

$$\hat{\boldsymbol{\alpha}}_{\ell_1 CLS} = \underset{\boldsymbol{\alpha}}{\mathrm{argmin}}\, J_{\ell_1 CLS}(\boldsymbol{\alpha})$$

- It is often called ℓ_1 regularized LS.

How to Obtain Solutions (cont.)

- How to deal with ℓ_1-norm?
- Use the following identity:

$$\|\boldsymbol{\alpha}\|_1 = \min_{\boldsymbol{u}\in\mathbb{R}^p} \sum_{i=1}^{p} u_i \quad \text{subject to} \quad -\boldsymbol{u} \leq \boldsymbol{\alpha} \leq \boldsymbol{u},$$

- $\hat{\boldsymbol{\alpha}}_{\ell_1 CLS}$ is given by the solution of

$$\min_{\boldsymbol{\alpha},\boldsymbol{u}} \left[J_{LS}(\boldsymbol{\alpha}) + \lambda \sum_{i=1}^{p} u_i \right] \quad \text{subject to} \quad -\boldsymbol{u} \leq \boldsymbol{\alpha} \leq \boldsymbol{u},$$

Linearly Constrained Quadratic Programming Problem

- Standard optimization softwares can solve the following form of linearly constrained quadratic programming problems.

$$\min_{\boldsymbol{\beta}} \left[\frac{1}{2} \langle \boldsymbol{Q}\boldsymbol{\beta}, \boldsymbol{\beta} \rangle + \langle \boldsymbol{\beta}, \boldsymbol{q} \rangle \right]$$

$$\text{subject to } \boldsymbol{V}\boldsymbol{\beta} \leq \boldsymbol{v}$$

$$\boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g}$$