Optimization methods and game theory

- Karush-Kuhn-Tucker conditions
- Support Vector Machines
- ε-SV Regression
- Clustering
- Solution methods
- Multiobjective optimization
- Game theory

KKT

Given the problem

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0 \forall i \\ h_i(x) = 0 \forall i \end{cases} \qquad (P)$$

Abadie Constraint Qualification

ACQ holds if $T_x(x^*) = D(x^*)$

Sufficient conditions

- Affine constraints
- Slater's condition

  if all the g_i are convex and all the h_i are affine and there exists an interior point in the feasible set $\bar{x} : g_i(\bar{x}) < 0 \forall i$ and $h_i(\bar{x}) = 0 \forall i$
- Linear independence of the gradients of active constraints

KKT System

One of the solutions to the KKT system is the optimal solution to the problem

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla h_i(x^*) = 0 \\ g_i(x^*) \leq 0 \forall i \\ h_i(x^*) = 0 \forall i \\ \lambda_i \geq 0 \forall i \\ \lambda_i g_i(x^*) = 0 \forall i \end{cases}$$

If the problem is convex the solution is a global optimum.

The inf $L(x, \lambda, \mu)$ is the Lagrangian relaxation of the problem P, provides a lower bound to the optimal value of P.

$\varphi(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$ is the dual function.

This is the dual problem

$$\begin{cases} \max \varphi(\lambda, \mu) \\ \lambda \geq 0 \end{cases} \qquad (D)$$

$v(D) \leq v(P)$

The dual problem is always a convex optimization problem even if the primal problem is not.

If the primal is continuously differentiable and convex and ACQ holds at the solution, then v(D)=v(P).

$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$ iff x^*, λ^*, μ^* is an optimal solution and strong duality holds.

Support Vector Machines

Linear SVM

We must solve the problem

$$\begin{cases} \min \frac{1}{2}||w||^2 \\ 1 - y_i(w^T x_i + b) \leq 0 \forall i \end{cases}$$

The dual problem is

$$\begin{cases} \max_\lambda -\frac{1}{2} \sum_i \sum_j y^i y^j (x^i)^T x^j \lambda_i \lambda_j + \sum_i \lambda_i \\ \sum_i y^i \lambda_i = 0 \\ \lambda_i \geq 0 \forall i \end{cases}$$

Also written as

$$\begin{cases} \max_\lambda -\frac{1}{2} \lambda^T X^T X \lambda + 1^T \lambda \\ y^T \lambda = 0 \\ \lambda \geq 0 \end{cases}$$

Where $X = \{y^i x^i\}$.

Linear SVM with slack variables

We must solve the problem

$$\begin{cases} \min \frac{1}{2}||w||^2 + C\sum_i \xi_i \\ 1 - y_i(w^T x_i + b) \leq -\xi_i \forall i \\ \xi_i \geq 0 \forall i \end{cases}$$

The dual problem is

$$\begin{cases} \max_\lambda -\frac{1}{2}\sum_i \sum_j y^i y^j (x^i)^T x^j \lambda_i \lambda_j + \sum_i \lambda_i \\ \sum_i y^i \lambda_i = 0 \\ 0 \leq \lambda_i \leq C \forall i \end{cases}$$

Also written as

$$\begin{cases} \max_\lambda -\frac{1}{2}\lambda^T X^T X \lambda + 1^T \lambda \\ y^T \lambda = 0 \\ 0 \leq \lambda \leq C \end{cases}$$

Where $X = \{y^i x^i\}$.

Kernel SVM with slack variables

We must solve the problem

$$\begin{cases} \min \frac{1}{2}||w||^2 + C\sum_i \xi_i \\ 1 - y_i(w^T \phi(x_i) + b) \leq -\xi_i \forall i \\ \xi_i \geq 0 \forall i \end{cases}$$

w might be infinite-dimensional so we have to use the dual problem.

The dual problem is

$$\begin{cases} \max_\lambda -\frac{1}{2}\sum_i \sum_j y^i y^j \phi(x^i)^T \phi(x^j) \lambda_i \lambda_j + \sum_i \lambda_i \\ \sum_i y^i \lambda_i = 0 \\ 0 \leq \lambda_i \leq C \forall i \end{cases}$$

Also written as

$$\begin{cases} \max_\lambda -\frac{1}{2}\lambda^T K \lambda + 1^T \lambda \\ y^T \lambda = 0 \\ 0 \leq \lambda \leq C \end{cases}$$

Where $K = \{k_{ij} = y^i y^j k(x^i, x^j)\}$ is the kernel matrix and $k(x^i, x^j) = \phi(x^i)^T \phi(x^j)$ is the kernel function.

In this way we never need to compute $\phi(x)$.

We then choose an i s.t. $0 < \lambda_i < C$ and compute $b^* = \frac{1}{y^i} - \sum_j \lambda_j^* y^j k(x^i, x^j)$.

The decision function will be $f(x) = \sum_i \lambda_i^* y^i k(x^i, x) + b^*$.

ε-SV Regression

Linear ε-SV Regression

The problem is

$$\begin{cases} \min \frac{1}{2}||w||^2 \\ |y_i - w^T x_i - b| \leq \varepsilon \forall i \end{cases}$$

We can split the absolute value into two inequalities:

$$\begin{cases} \min \frac{1}{2}||w||^2 \\ y_i - w^T x_i - b \leq \varepsilon \forall i \\ -y_i + w^T x_i + b \leq \varepsilon \forall i \end{cases}$$

Linear ε-SV Regression with slack variables

The problem is

$$\begin{cases} \min \frac{1}{2}||w||^2 + C \sum_i \xi_i \\ y_i - w^T x_i - b \leq \varepsilon + \xi_i^+ \forall i \\ -y_i + w^T x_i + b \leq \varepsilon + \xi_i^- \forall i \\ \xi_i^+ \geq 0 \forall i \\ \xi_i^- \geq 0 \forall i \end{cases}$$

The dual problem is

$$\begin{cases} \max_{\lambda^+, \lambda^-} -\frac{1}{2} \sum_i \sum_j (\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)(x^i)^T x^j \\ \qquad\qquad -\varepsilon \sum_i (\lambda_i^+ + \lambda_i^-) + \sum_i y_i (\lambda_i^+ - \lambda_i^-) \\ \sum_i (\lambda_i^+ - \lambda_i^-) = 0 \\ 0 \leq \lambda_i^+ \leq C \forall i \\ 0 \leq \lambda_i^- \leq C \forall i \end{cases}$$

In matrix form

$$
\begin{cases}
\max_{\lambda^+,\lambda^-} -\frac{1}{2}\begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix}^T Q \begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix} + \left(-\varepsilon 1^T + \begin{pmatrix} y\\ -y\end{pmatrix}^T\right)\begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix}\\
1^T(\lambda^+ - \lambda^-) = 0\\
0 \leq \lambda^+ \leq C\\
0 \leq \lambda^- \leq C
\end{cases}
$$

With $Q = \begin{pmatrix} K & -K\\ -K & K\end{pmatrix}$ and $K_{ij} = x^i \cdot x^j$.

We can find w and b from the dual solution

$$
w = \sum_i (\lambda_i^+ - \lambda_i^-) x_i
$$

If $\exists i$ s.t. $0 < \lambda_i^+ < C$ then $b = y_i - w^T x_i - \varepsilon$.

If $\exists i$ s.t. $0 < \lambda_i^- < C$ then $b = y_i - w^T x_i + \varepsilon$.

Non-linear ε-SV Regression

We can just replace $x^i \cdot x^j$ with $K(x^i, x^j)$ in the dual problem of the linear ε-SV Regression with slack variables.

The regression function is

$$
f(x) = \sum_i (\lambda_i^+ - \lambda_i^-) K(x^i, x) + b
$$

Where b is computed as

If $\exists i$ s.t. $0 < \lambda_i^+ < C$ then

$$
b = y_i - \varepsilon - \sum_j (\lambda_j^+ - \lambda_j^-) K(x^j, x^i)
$$

If $\exists i$ s.t. $0 < \lambda_i^- < C$ then

$$
b = y_i + \varepsilon - \sum_j (\lambda_j^+ - \lambda_j^-) K(x^j, x^i)
$$

Clustering

K-means

The problem is

$$\begin{cases} \min \sum_i \min_{j=1\dots k} ||p_i - x_j||_2^2 \\ x_j \in \mathbb{R}^n \forall j \end{cases}$$

If $k = 1$ the solution is $x_1 = \frac{1}{n} \sum_i p_i$.

If $k > 1$ the problem is non-convex and non differentiable.

If we fix p_i and x_j then the problem is

$$\begin{cases} \min \sum_{j=1\dots k} \alpha_{ij} ||p_i - x_j||_2^2 \\ \sum_{j=1\dots k} \alpha_{ij} = 1 \\ \alpha_{ij} \geq 0 \forall j \end{cases}$$

An optimal solution is $\alpha_{ij} = 1$ if $j = \text{argmin}_{j=1\dots k} ||p_i - x_j||_2^2$ and 0 otherwise.

The problem is equivalent to

$$\begin{cases} \min \sum_i \sum_{j=1\dots k} \alpha_{ij} ||p_i - x_j||_2^2 \\ \sum_{j=1\dots k} \alpha_{ij} = 1 \\ \alpha_{ij} \geq 0 \forall j \\ x_j \in \mathbb{R}^n \forall j \end{cases}$$

If we fix x_j then the problem is decomposable in n simple LP problems (in α_{ij}).

$$\alpha_{ij}^* = \begin{cases} 1 & \text{if } j = \text{argmin}_{j=1\dots k} ||p_i - x_j||_2^2 \\ 0 & \text{otherwise} \end{cases}$$

If we fix α_{ij} then the problem is decomposable in k convex QP problems in x.

$$x_j^* = \frac{\sum_i \alpha_{ij} p_i}{\sum_i \alpha_{ij}}$$

We can then create an algorithm of alternating minimization:

1. Initialize x_j randomly and assign α_{ij} as above.
2. Update x_j as above.
3. Update α_{ij} as above.
4. Given $f(x, \alpha) = \sum_i \sum_j \alpha_{ij} ||p_i - x_j||_2^2$
   - If $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ then STOP.

- Else go to Step 2.

K-median

If we replace the L_2 norm with the L_1 norm we get the K-median problem.

The solution is equivalent to the K-means problem but we update x_j with the median instead of the mean.

Solution methods

Linear equality constraints

The problem

$$\begin{cases} \min f(x) \\ Ax = b \end{cases}$$

A can be written as $A = [A_1, A_2]$ with $det(A_1) \neq 0$ and $A_1 \in \mathbb{R}^{p \times p}$. x can be written as $x = [x_1, x_2]$ with $x_1 \in \mathbb{R}^p$

We can then set $x_1 = A_1^{-1}(b - A_2 x_2)$ and thus eliminating the variables x_1 from the problem.

The problem becomes unconstrained in $n - p$ variables x_2

$$\begin{cases} \min f(A_1^{-1}(b - A_2 x_2), x_2) \\ x_2 \in \mathbb{R}^{n-p} \end{cases}$$

Penalty method

The problem

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0 \forall i \end{cases} \qquad (P)$$

With X the feasible set.

We can define the penalty function

$$p(x) = \sum_{i=1}^{m} \max(0, g_i(x))$$

And then the problem becomes

$$\begin{cases} \min f(x) + \frac{1}{\varepsilon} p(x) := f_\varepsilon(x) \\ x \in \mathbb{R}^n \end{cases} \qquad (P_\varepsilon)$$

If x^* solves (P_ε) and $x^* \in X$ then x^* also solves (P).

The algorithm to find the solution of (P) is

1. Set $\varepsilon = \varepsilon_0$ and $\tau \in (0, 1)$.
2. Solve (P_ε) and get x^*.
3. Then
   - If $x^* \in X$ then STOP.
   - Else set $\varepsilon = \tau\varepsilon$ and go to step 2.

Logarithmic barrier method

The problem

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0 \forall i \end{cases} \qquad (P)$$

With X the feasible set.

Can be approximated inside $\text{int}(X)$ by the problem

$$\begin{cases} \min f(x) - \varepsilon \sum_{i=1}^m \log(-g_i(x)) := \psi_\varepsilon(x) \\ x \in \text{int}(X) \end{cases} \qquad (P_\varepsilon)$$

We call $B(x) = -\sum_{i=1}^m \log(-g_i(x))$ the barrier function. So $\psi_\varepsilon(x) = f(x) - \varepsilon B(x)$.

Note that as x approaches the boundary of X, $\psi_\varepsilon(x) \to +\infty$

If x^* is a local minimum of (P_ε) then

$$\nabla \psi_\varepsilon(x^*) = \nabla f(x^*) - \varepsilon \sum_{i=1}^m \frac{\nabla g_i(x^*)}{-g_i(x^*)} = 0$$

We can show that $v(P) = v(P_\varepsilon) - m\varepsilon$ where m is the number of constraints.

The algorithm is

1. Set the tolerance $\delta > 0$ and $\tau \in (0, 1)$ and $\varepsilon_1 > 0$. Choose $x^0 \in \text{int}X$ set $k = 1$

2. Find the optimal solution x^k of

$$\begin{cases} \min \psi_\varepsilon(x) \\ x \in \text{int}X \end{cases}$$

using x^{k-1} as a starting point.

3. Then
   - If $m\varepsilon_k < \delta$ then STOP.
   - Else $\varepsilon_{k+1} = \tau\varepsilon_k$ and $k = k+1$ and go to step 2.

To find a starting point x^0 we can solve the problem

$$\begin{cases} \min s \\ g_i(x) \leq s \forall i \end{cases}$$

With the Logarithmic barrier method starting from any $\tilde{x} \in \mathbb{R}^n$ and $\tilde{s} > \max g(\tilde{x})$. If $s^* < 0$ then $x^* \in \text{int}X$, otherwise $\text{int}X = \emptyset$.

Multiobjective optimization

The problem is defined as follows:

$$\begin{cases} \min f(x) = (f_1(x), f_2(x), \ldots, f_s(x)) \\ x \in X \end{cases}$$

Given $x, y \in \mathbb{R}^s$ we say that $x \geq y$ if $x_i \geq y_i \forall i = 1, \ldots, s$.

- A point $\bar{x}$ is said to be Pareto ideal minimum if $\bar{x} \leq x \forall x \in X$
- A point $\bar{x}$ is said to be Pareto minimum if $\nexists x \in X : x \neq \bar{x}$ and $\bar{x} \geq x$
- A point $\bar{x}$ is said to be Pareto weak minimum if $\nexists x \in X : \bar{x} > x$ and $\bar{x}_i > x_i \forall i$

Auxiliary optimization problem

$x^* \in X$ is a minimum of (P) iff the auxiliary optimization problem

$$\begin{cases} \max \sum_{i=1}^{s} \varepsilon_i \\ f_i(x) + \varepsilon_i \leq f_i(x^*) \forall i \\ x \in X \\ \varepsilon \geq 0 \end{cases}$$

has optimal value 0

$x^* \in X$ is a weak minimum of (P) iff the auxiliary optimization problem

$$\begin{cases} \max v \\ v \leq \varepsilon_i \forall i \\ f_i(x) + \varepsilon_i \leq f_i(x^*) \forall i \\ x \in X \\ \varepsilon \geq 0 \\ \sum_{i=1}^s \varepsilon_i = 0 \end{cases}$$

has optimal value 0

If x^* is a weak minimum then there exists $\theta^* \in \mathbb{R}^s$ such that (x^*, θ^*) is a solution of the system

$$\begin{cases} \sum_{i=1}^s \theta_i \nabla f_i(x) = 0 \\ \theta_i \geq 0 \forall i \\ \sum_{i=1}^s \theta_i = 1 \\ x \in \mathbb{R}^n \end{cases} \qquad (S)$$

If the problem is convex, the above condition is also sufficient. If $\theta^* > 0$ then x^* is a minimum.

If x^* is a weak minimum of (P) and ACQ holds at x^*, then there exists $\theta^* \in \mathbb{R}^s, \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p$ such that $(x^*, \theta^*, \lambda^*, \mu^*)$ is a solution of the system

$$\begin{cases} \sum_{i=1}^s \theta_i \nabla f_i(x) + \sum_{j=1}^m \lambda_j \nabla g_j(x) + \sum_{k=1}^p \mu_k \nabla h_k(x) = 0 \\ \theta_i \geq 0 \forall i \\ \sum_{i=1}^s \theta_i = 1 \\ \lambda \geq 0 \\ \lambda_j g_j(x^*) = 0 \forall j \\ g_j(x) \leq 0, h_k(x) = 0 \end{cases}$$

If the problem is unconstrained then the KKT system reduces to (S) If $\theta^* > 0$ then x^* is a minimum.

Weighted sum method

Given the problem

$$\begin{cases} \min f(x) = (f_1(x), f_2(x), \dots, f_s(x)) \\ x \in X \end{cases} \qquad (P)$$

And a set of weights $\alpha = \{\alpha_1, \dots \alpha_s\} \geq 0$ associated with the objectives f_i.

We associate with (P) the scalar problem

$$\begin{cases} \min \sum_{i=1}^{s} \alpha_i f_i(x) \\ x \in X \end{cases} \qquad (P_\alpha)$$

The solutions of (P_α) are weak minima of (P) if $\alpha_i \geq 0$ for all i and are minima if $\alpha_i > 0$ for all i.

If the problem is convex, any weak minimum of (P) can be obtained given the right weights α.

If (P) is linear and X is a polyhedron, then any minimum of (P) can be obtained given the right weights α.

Goal method

Define $z_i = \min_{x \in X} f_i(x) \forall i$

We want to find the closest point to z in $f(X)$.

$$\begin{cases} \min \sum_{i=1}^{s} ||f_i(x) - z_i||_q \\ x \in X \end{cases} \qquad (G)$$

If $q \in [1, +\infty)$ then any optimal solution of (G) is a minimum of (P).

If $q = +\infty$ then any optimal solution of (G) is a weak minimum of (P).

Game theory

A non-cooperative game is defined by a set of N players, each player i with a set of strategies X_i and a cost function $f_i : X_1 \times \ldots \times X_N \to \mathbb{R}$. The aim of each player is to solve

$$\begin{cases} \min f_i(x_1, \ldots, x_N) \\ x_i \in X_i \end{cases}$$

Nash equilibrium

In a two-player non-cooperative game (2PNCG), a Nash equilibrium is a pair $(\bar{x}, \bar{y})$ s.t.

$$f_1(\bar{x}, \bar{y}) = \min_{x \in X} f_1(x, \bar{y})$$

and

$$f_2(\bar{x}, \bar{y}) = \min_{y \in Y} f_2(\bar{x}, y)$$

Matrix game

A matrix game is a 2PNCG with two finite sets of strategies $X = \{1, \ldots, m\}$ and $Y = \{1, \ldots, n\}$ and $f_2 = -f_1$. The game can be represented by a matrix $A = \{a_{ij} = f_1(i, j)\}$. (Player 2 wants to maximize the value on the matrix, while player 1 wants to minimize it.)

To have a Nash equilibrium $(\bar{x}, \bar{y})$ we need to have

$$f_1(\bar{x}, \bar{y}) = \min_{x \in X} f_1(x, \bar{y})$$

and

$$f_1(\bar{x}, \bar{y}) = \max_{y \in Y} f_1(\bar{x}, y)$$

To find a Nash equilibrium, we can remove rows and columns that are strictly dominated by other rows and columns until we have reduced the matrix to a 1×1 matrix.

A strategy is strictly dominated if there is another strategy that is always better than it.

Mixed strategies In a matrix game C, a mixed strategy is a vector of probabilities $x = (x_1, \ldots, x_m)$ s.t. $x_i \geq 0$ and $\sum_{i=1}^{m} x_i = 1$ for player 1 and $y = (y_1, \ldots, y_n)$ s.t. $y_j \geq 0$ and $\sum_{j=1}^{n} y_j = 1$ for player 2.

The expected value of the game is $f_1(x, y) = x^T C y$ and $f_2(x, y) = -x^T C y$.

A Nash equilibrium exists if

$$\max_{y \in Y} \bar{x}^T C y = \bar{x}^T C \bar{y} = \min_{x \in X} x^T C \bar{y}$$

$(\bar{x}, \bar{y})$ is a saddle point of $f_1(x, y)$.

It is always possible to find a mixed strategy Nash equilibrium in a finite matrix game.

The problem

$$\min_{x \in X} \max_{y \in Y} f_1(x, y)$$

is equivalent to

$$\begin{cases} \min v \\ v \geq \sum_{i=1}^{m} c_{ij} x_i \forall j \\ x \geq 0 \\ \sum_{i=1}^{m} x_i = 1 \end{cases} \qquad (P_1)$$

and

$$\max_{y \in Y} \min_{x \in X} f_1(x, y)$$

is equivalent to

$$\begin{cases} \max w \\ w \leq \sum_{j=1}^{n} c_{ij} y_j \forall i \\ y \geq 0 \\ \sum_{j=1}^{n} y_j = 1 \end{cases} \qquad (P_2)$$

These two problems are one the dual of the other.

Bimatrix game

A bimatrix game is a matrix game but with two matrices, one for each player. This means that $f_2 \neq -f_1$. $f_1(x, y) = x^T C_1 y$ and $f_2(x, y) = x^T C_2 y$.

Any bimatrix game has a mixed strategy Nash equilibrium.

$(\bar{x}, \bar{y})$ is a Nash equilibrium iff $\exists \mu_1, \mu_2$ s.t.

$$\begin{cases} C_1 \bar{y} + \mu_1 1_m \geq 0 \\ \bar{x} \geq 0 \\ \sum_{i=1}^{m} x_i = 1 \\ \bar{x}_i (C_1 \bar{y} + \mu_1 1_m)_i = 0 \forall i \\ C_2 \bar{x} + \mu_2 1_n \geq 0 \\ \bar{y} \geq 0 \\ \sum_{j=1}^{n} y_j = 1 \\ \bar{y}_j (C_2 \bar{x} + \mu_2 1_n)_j = 0 \forall j \end{cases} \qquad (KS)$$

This is the KKT system for the problem.

$(\bar{x}, \bar{y}, \mu_1, \mu_2)$ is a solution of the KKT system iff it's an optimal solution of this quadratic programming problem:

$$
\begin{cases}
\min \psi(x, y, \mu_1, \mu_2) = \\
\qquad x^T(C_1 y + \mu_1 1_m) + y^T(C_2 x + \mu_2 1_n) \\
C_1 y + \mu_1 1_m \geq 0 \\
x \geq 0 \\
\sum_{i=1}^m x_i = 1 \\
C_2^T x + \mu_2 1_n \geq 0 \\
y \geq 0 \\
\sum_{j=1}^n y_j = 1
\end{cases} \tag{QP}
$$

Convex game

Consider a general 2PNCG

$$
\begin{cases}
\min f_1(x, y) \\
g_i^1(x) \leq 0 \forall i
\end{cases}
\qquad
\begin{cases}
\min f_2(x, y) \\
g_j^2(y) \leq 0 \forall j
\end{cases}
$$

If X and Y are convex sets, $f_1(\cdot, y)$ is quasiconvex $\forall y \in Y$ and $f_2(x, \cdot)$ is quasiconvex $\forall x \in X$, then the game has a Nash equilibrium.

If $(\bar{x}, \bar{y})$ is a Nash equilibrium and ACQ holds, then the double KKT system is satisfied:

$$
\begin{cases}
\nabla_x f_1(\bar{x}, \bar{y}) + \sum_{i=1}^p \lambda_i^1 \nabla g_i^1(\bar{x}) = 0 \\
\lambda_i^1 \geq 0 \\
g^1(\bar{x}) \leq 0 \\
\lambda_i^1 g_i^1(\bar{x}) = 0 \forall i \\
\nabla_y f_2(\bar{x}, \bar{y}) + \sum_{j=1}^q \lambda_j^2 \nabla g_j^2(\bar{y}) = 0 \\
\lambda_j^2 \geq 0 \\
g^2(\bar{y}) \leq 0 \\
\lambda_j^2 g_j^2(\bar{y}) = 0 \forall j
\end{cases}
$$