A polynomial case of cardinality constrained quadratic optimization problem*

Jianjun Gao † Duan Li ‡

August 27, 2010

Abstract

We investigate in this paper a fixed parameter polynomial algorithm for the cardinality constrained quadratic optimization problem, which is NP-hard in general. More specifically, we prove that, given a problem of size n, the number of decision variables, and s, the cardinality, if, for some $0 < k \leq n$, the $n-k$ largest eigenvalues of the coefficient matrix of the problem are identical, we can construct a solution algorithm with computational complexity of $\mathcal{O}\left(n^{2k}\right)$, which is independent of the cardinality s. Our main idea is to decompose the primary problem into several convex subproblems, while the total number of the subproblems is determined by the cell enumeration algorithm for hyperplane arrangement in $\mathbb{R}^k$ space.

Keywords: Cardinality constrained quadratic optimization, cell enumeration, nonconvex optimization, fixed parameter polynomial algorithm

1 Introduction

We consider in this paper the following cardinality constrained quadratic optimization problem (CCQO),

$$
\begin{aligned}
(\mathcal{P}): \quad & \min_{x} \quad f(x) = \frac{1}{2}x'Qx + q'x \\
\text{Subject to:} \quad & x \in \Delta(s) \triangleq \left\{ x \in \mathbb{R}^n \mid \sum_{t=1}^{n} \delta(x_i) \leq s < T \right\},
\end{aligned} \tag{1}
$$

where $Q \in \mathbb{R}^{n\times n}$ is positive definite, $q \in \mathbb{R}^n \setminus \{0\}$, and the indicator function $\delta(\cdot): \mathbb{R} \to \{0,1\}$ is defined such that $\delta(a) = 1$ if a is non-zero and $\delta(a) = 0$ otherwise.

*This work was supported by Research Grants Council of Hong Kong, under grants 414207 and 414808.

†Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong (jjgao@se.cuhk.edu.hk)

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong.(dli@se.cuhk.edu.hk).

This class of quadratic optimization problems with a cardinality constraint arises naturally from various applicationsThe exact and approximate solution approaches for cardinality constrained optimization problems are studied in the literatures, e.g., [2], [3], [9], [12]. To our best knowledge, only a few results on polynomially solvable cases of CCQO problem are reported in the literature. In the context of subset selection problem, Das and Kempe [5] proposed an approximate algorithm for the case where the covariance possesses a constant bandwidth and developed an exact algorithm for the case where the covariance graph is of a tree structure. Recently, Donoho and Candes [4][6] showed that, under some conditions, using l_1 norm to replace the cardinality constraint in the sparse signal reconstruction problem yields the exact solution with an overwhelming probability.

We focus in this paper on a class of CCQO problems with a special structure. More specifically, we consider situations where the $n-k$ largest eigenvalues of matrix Q are identical, $0 \leq k \leq n$, which we term as a matrix Q with a k-degree freedom (see Definition 2.1). We prove that, for fixed k, this class of CCQO problems is polynomially solvable. Motivated by the geometrical characteristics of CCQO problem, we decompose the problem $(\mathcal{P})$ into several convex quadratic programming subproblems and the number of these sub-problems is determined by a cell enumeration algorithm for the hyperplane arrangement in $\mathbb{R}^k$ space [1]. From the complexity point of view, if k is fixed, the solution scheme is a polynomial-time algorithm. To certain extent, our result in this paper is similar to a polynomially solvable case in binary quadratic program, where the rank of coefficient matrix is fixed (see [8]).

This paper is organized as following. After the introduction in this section, we develop the solution scheme for CCQO problem with a k-degree freedom coefficient matrix in Section 2. As the derived solution scheme depends heavily on a distance function between the cardinality feasible set and an affine space, we develop a scheme for identifying such a distance function in Section 3 using cell enumeration of hyperplane arrangement in discrete geometry. After presenting an illustrative example in Section 4, we conclude the paper in Section 5.

Throughout the paper, we use $v(\cdot)$ to denote the optimal value of problem $(\cdot)$, $S \succ 0$ a positive definite matrix, $\mathbb{S}^n_{++}$ the set of positive definite matrices, $\mathrm{diag}\{a\} \in \mathbb{R}^{n\times n}$ the diagonal matrix with $a \in \mathbb{R}^n$ being its diagonal, $\mathbf{0}$ the vector with all elements being 0 and $\|\cdot\|$ the l_2 norm. Furthermore, we denote the ellipsoid and the ball in $\mathbb{R}^n$, respectively, by

$$\mathcal{E}(P,p,\rho) \triangleq \left\{y \in \mathbb{R}^n \mid (y-p)'P(y-p) \leq \rho\right\},\ P \succ 0, \rho \geq 0, \tag{2}$$

$$\mathcal{B}(p,r^2) \triangleq \left\{y \in \mathbb{R}^n \mid \|y-p\|^2 \leq r^2\right\}. \tag{3}$$

2 Solution scheme of problem $(\mathcal{P})$

2.1 Preliminary

Problem $(\mathcal{P})$ has been proved to be, in general, NP-hard (see the proof in [11]). Here we give an alternative proof for the NP-hardness of problem $(\mathcal{P})$, which appears to be much simpler than the one in [11]. Let us construct the following problem,

$$(\mathcal{G}):\ \min_{x\in\mathbb{R}^n}\ \left\{\hat{f} := M\|x-\mathbf{1}\|^2 + \|Ax\|^2 \mid x \in \Delta(s)\right\},$$

where $M > 0$ is a large number, $\mathbf{1}$ is the vector with all elements being 1, and $A \in \mathbb{R}^{l\times n}$ with $l \le n$. Note that any instance of problem $(\mathcal{G})$ is polynomially reducible to an instance of problem $(\mathcal{P})$. That is to say, solving problem $(\mathcal{G})$ is no more difficult than solving problem $(\mathcal{P})$. Since $x \in \Delta(s)$, minimizing the first term of $(\mathcal{G})$ enforces x_i to take either 0 or 1 for $i = 1, \cdots, n$. More specifically, at least $n-s$ of x_i's are zero. Thus, the optimal value of problem $\mathcal{G}$ is lower bounded, i.e., $v(\mathcal{G}) \ge M(n-s)$. Answering the question "whether equality $v(\mathcal{G}) = M(n-s)$ holds or not " turns out to find the integer (binary) solution of linear systems $Ax = 0$ such that $x \in \{0,1\}^n$ and $\sum_{i=1} x_i \le s$, which is a known NP-complete decision problem [7]. Our conclusion for the NP-hardness of problem $(\mathcal{P})$ follows the simple reduction method [7].

Geometrically, the objective contour of $(\mathcal{P})$ is an ellipsoid in $\mathbb{R}^n$ space,

$$\begin{aligned}\mathcal{E}(Q,h,\rho) &\triangleq \{x \in \mathbb{R}^n \mid f(x) \le \tau\}\\ &= \{x \in \mathbb{R}^n \mid (x-h)'Q(x-h) \le \rho\},\end{aligned}$$

where

$$h \triangleq -Q^{-1}q, \tag{4}$$

$$C \triangleq -\frac{1}{2}q'Q^{-1}q, \tag{5}$$

$$\rho \triangleq 2\tau - 2C. \tag{6}$$

Clearly, we must have $\tau \ge C$. Minimizing $f(x)$ under constraint (1) is now equivalent to finding the minimum ellipsoid that touches the set $\Delta(s)$, or equivalently, we can reformulate problem $(\mathcal{P})$ as follows,

$$(\mathcal{P}_1): \quad \min_{x,\rho} \quad \frac{1}{2}\rho + C, \tag{7}$$

$$\text{Subject to:} \quad x \in \mathcal{E}(Q,h,\rho), \tag{8}$$

$$x \in \Delta(s).$$

In the following, we choose to deal with problem formulation $(\mathcal{P}_1)$, instead of problem formulation $(\mathcal{P})$. Figure 1 illustrates a case where $n = 2$ and $s = 1$ and the feasible set $\Delta(s)$ consists of both x-axis and y-axis in $\mathbb{R}^2$ plane. It is clear from the figure that the optimal contour is the minimum ellipsoid that touches the x-axis. Note that the number of feasible subspaces could be as large as $\sum_{j=1}^{s} \mathcal{C}_n^j$, where $\mathcal{C}_n^j = \frac{n!}{j!(n-j)!}$.

Let the spectral decomposition of matrix Q be $Q = \Gamma'\Lambda^Q\Gamma$, where matrix Γ is unitary and

$$\Lambda^Q \triangleq \mathrm{diag}\{\lambda_1^Q, \lambda_2^Q, \cdots, \lambda_n^Q\}. \tag{9}$$

Without loss of generality, we assume the eigenvalues of matrix Q to be arranged in an ascending order,

$$0 < \lambda_1^Q \le \lambda_2^Q \le \cdots \le \lambda_n^Q.$$

Definition 2.1. Matrix $H \in \mathbb{S}_+^n$ is said to be of k-degree freedom, if there exists k such that $0 \le k \le n$ and $0 \le \lambda_1^H \le \lambda_2^H \cdots \lambda_k^H < \lambda_{k+1}^H = \lambda_{k+2}^H = \cdots = \lambda_n^H$, where λ_i^H is the i-th smallest eigenvalue of H.

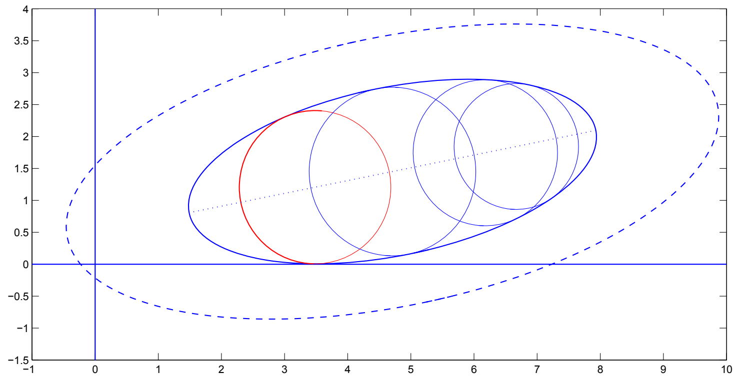

Figure 1: The optimal contour and $\Delta(s)$ when $n = 2$ and $s = 1$

In plain words, a positive definite matrix is of k-degree freedom, if the $n-k$ largest eigenvalues are identical. If $k = 0$, the 0-degree freedom matrix is a diagonal matrix with identical eigenvalues. If $k = n$, all the eigenvalues of the matrix have their full freedom. It is very interesting to investigate the cases when $0 < k < n$, as we will demonstrate in the following that any CCQO problems can be always approximated by such a matrix of k-degree freedom. For problem $(\mathcal{P})$, we can chose a k and construct an auxiliary problem ,

$$\begin{aligned} (\mathcal{A}): \quad & \min_x \ \hat{f}(x) = \frac{1}{2}(x-h)'A(x-h) + C, \\ \text{Subject to:} \ & x \in \Delta(s), \end{aligned}$$

where $A \in \mathbb{S}^n_+$ is a matrix of k-degree freedom specified by $A = \Gamma'\Lambda_k\Gamma$ with

$$\Lambda_k = \text{diag}\big(\lambda^Q_1, \lambda^Q_2, \cdots, \lambda^Q_k, \lambda^Q_k, \cdots, \lambda^Q_k\big).$$

Lemma 2.1. *If $Q \succ 0$, the following relationships hold,*

$$v(\mathcal{A}) \le v(\mathcal{P}) \text{ and } v(\mathcal{P}) - v(\mathcal{A}) \le (\lambda^Q_n - \lambda^Q_k)\Phi,$$

where Φ is a parameter dependent on problem $(\mathcal{P})$.

Proof. Since $Q \succeq A$, we have $\hat{f}(x) \le f(x)$ for all $x \in \mathbb{R}^n$, which implies that $v(\mathcal{A}) \le v(\mathcal{P})$. Let $\hat{x}$ be the optimal solution of problem $(\mathcal{A})$, thus giving rise to

$$\begin{aligned} v(\mathcal{P}) - v(\mathcal{A}) \le f(\hat{x}) - \hat{f}(\hat{x}) &= \frac{1}{2}(\hat{x}-h)'(Q-A)(\hat{x}-h) \\ &\le \frac{1}{2}(\lambda^Q_n - \lambda^Q_k)\|\hat{x}-h\|^2. \end{aligned} \tag{10}$$

As the following holds true for any $x \in \Delta(x)$ in problem $(\mathcal{A})$,

$$\lambda^Q_1\|\hat{x}-h\|^2 + C \le \hat{f}(\hat{x}) \le \lambda^Q_k\|x-h\|^2 + C, \quad \forall x \in \Delta(x). \tag{11}$$

We can minimize the upper bound $\lambda_k^Q \|x-h\|^2$ in (11) by taking $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_n) \in \Delta(s)$ with $\bar{x}_i = h_i$ for $i \in \mathcal{I}$ and $\bar{x}_i = 0$ for $i \notin \mathcal{I}$, where $\mathcal{I}$ is the index set consisting of the first s largest elements of $|h_i|$. Then, inequality (11) becomes

$$\|\hat{x} - h\|^2 \leq \frac{\lambda_k^Q}{\lambda_1^Q} \sum_{i \notin \mathcal{I}} h_i^2. \tag{12}$$

Combining (10) and (12) yields the conclusion in the lemma. □

Lemma 2.1 suggests that increasing the order of k may reduce the gap between the auxiliary and the primary problems. However, the computational burden of solving auxiliary problem $(\mathcal{A})$ may increase for a larger k at the same time.

2.2 A decomposition approach

In this section, we develop an efficient solution scheme for problem $(\mathcal{P})$ with Q being of a k-degree freedom. Our main idea for solving such a class of problems is to decompose $(\mathcal{P})$ into several convex quadratic subproblems. Before we state our main results, we introduce some results on the decomposition of ellipsoid $\mathcal{E}(Q, h, \rho)$ when Q is of a k-degree freedom.

Theorem 2.1. *Let $\mathcal{E}(\Lambda, \mathbf{0}, \gamma)$ be an ellipsoid with $\gamma > 0$, where $\Lambda = diag\{\lambda_i\}|_{i=1}^n$ is of k-degree freedom with λ_i being the i-th smallest eigenvalue, $1 < k < n$.*

(i) For any $\alpha \in \mathbb{R}^n$ such that

$$\alpha \in E_k \triangleq \left\{ (u_1, \cdots, u_n)' \mid \sum_{i=1}^k \frac{\lambda_i \lambda_{k+1}^2 u_i^2}{(\lambda_{k+1} - \lambda_i)^2} \leq \gamma, \ \textit{and} \ \ u_j = 0, \ j = k+1, \cdots, n \right\},$$

the following holds,

$$r^2(\alpha) \triangleq \frac{\gamma}{\lambda_{k+1}} - \sum_{i=1}^k \frac{\lambda_i \alpha_i^2}{\lambda_{k+1} - \lambda_i} \geq 0. \tag{13}$$

(ii) The ellipsoid $\mathcal{E}(\Lambda, \mathbf{0}, \gamma)$ is the union of the balls expressed as follows,

$$\mathcal{E}(\Lambda, \mathbf{0}, \gamma) = \bigcup_{\alpha \in E_k} \mathcal{B}(\alpha, r^2(\alpha)).$$

Proof. (i) Since Λ is of k-degree freedom, we have $\lambda_{k+1} > \lambda_i > 0$ for $i = 1, \cdots, k$, which further implies $\lambda_{k+1}/(\lambda_{k+1} - \lambda_i) > 1$, for $i = 1, \cdots, k$. Then, for any $\alpha \in E_k$, the following inequality holds,

$$\sum_{i=1}^k \frac{\lambda_i \lambda_{k+1} \alpha_i^2}{(\lambda_{k+1} - \lambda_i)} \leq \sum_{i=1}^k \frac{\lambda_i \lambda_{k+1}^2 \alpha_i^2}{(\lambda_{k+1} - \lambda_i)^2} \leq \gamma.$$

Dividing both sides by λ_{k+1} gives rise to the result in (i).

(ii) For any $y^* \in \mathcal{E}(\Lambda, \mathbf{0}, \gamma)$, we have

$$\sum_{i=1}^{k} \lambda_i (y_i^*)^2 + \lambda_{k+1} \sum_{j=k+1}^{n} (y_j^*)^2 \leq \gamma. \tag{14}$$

Let $\alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*)'$ be defined such that $\alpha_i^* = y_i^*(\lambda_{k+1} - \lambda_i)/\lambda_{k+1}$ for $i = 1, \cdots, k$ and $\alpha_j^* = 0$ for $j = k+1, \cdots, n$. Then, we have

$$\sum_{i=1}^{k} \frac{\lambda_i \lambda_{k+1}^2 (\alpha_i^*)^2}{(\lambda_{k+1} - \lambda_i)^2} = \sum_{i=1}^{k} (y_i^*)^2 \lambda_i \leq \gamma - \lambda_{k+1} \sum_{i=k+1}^{n} (y_i^*)^2 \leq \gamma,$$

which implies that $\alpha^* \in E_k$. Define

$$r^2(\alpha^*) = \frac{\gamma}{\lambda_{k+1}} - \sum_{i=1}^{k} \frac{\lambda_i (\alpha_i^*)^2}{\lambda_{k+1} - \lambda_i}.$$

From the result in (i), we have $r^2(\alpha^*) \geq 0$. We can further conclude that $y^* \in \mathcal{B}(\alpha^*, r^2(\alpha^*))$ by checking the following inequality,

$$\begin{aligned} \|y^* - \alpha^*\|_2^2 - r^2(\alpha^*) = &\sum_{i=1}^{k} (y_i^* - y_i^* \frac{\lambda_{k+1} - \lambda_i}{\lambda_{k+1}})^2 + \sum_{j=k+1}^{n} (y_j^*)^2 \\ &- (\frac{\gamma}{\lambda_{k+1}} - \sum_{i=1}^{k} \frac{(y_i^*)^2 \lambda_i (\lambda_{k+1} - \lambda_i)}{\lambda_{k+1}^2}), \\ = &\sum_{i=1}^{k} \frac{(y_i^*)^2 \lambda_i}{\lambda_{k+1}} + \sum_{j=k+1}^{n} (y_j^*)^2 - (\frac{\gamma}{\lambda_{k+1}}) \leq 0, \end{aligned}$$

where the last inequality is implied by (14). Thus, we conclude that, for any $y^* \in \mathcal{E}(\Lambda, \mathbf{0}, \gamma)$, there exists $\alpha^* \in E_k$ such that $y^* \in \mathcal{B}(\alpha^*, r^2(\alpha^*))$, which further implies $\mathcal{E}(\Lambda, \mathbf{0}, \gamma) \subseteq \bigcup_{\alpha^* \in E_k} \mathcal{B}(\alpha^*, r^2(\alpha^*))$.

On the other hand, for any $\bar{y} \in \bigcup_{\alpha \in E_k} \mathcal{B}(\alpha, r^2(\alpha))$, there exists $\bar{\alpha} \in E_k$ such that $\bar{y} \in \mathcal{B}(\bar{\alpha}, r^2(\bar{\alpha}))$ with

$$r^2(\bar{\alpha}) = \frac{\gamma}{\lambda_{k+1}} - \sum_{i=1}^{k} \frac{\bar{\alpha}_i^2 \lambda_i}{\lambda_{k+1} - \lambda_i} \geq 0.$$

As, for $i = 1, \cdots, k$,

$$\frac{\lambda_{k+1} - \lambda_i}{\lambda_{k+1}} (\bar{y}_i - \frac{\bar{\alpha}_i \lambda_{k+1}}{\lambda_{k+1} - \lambda_i})^2 \geq 0,$$

we have

$$\frac{\bar{y}_i^2 \lambda_i}{\lambda_{k+1}} \leq (\bar{y}_i - \bar{\alpha}_i)^2 + \frac{\bar{\alpha}_i^2 \lambda_i}{\lambda_{k+1} - \lambda_i}, \quad i = 1, \cdots, k, \tag{15}$$

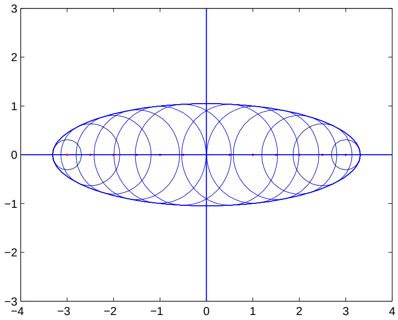

Figure 2: Decomposition of ellipsoid in $\mathbb{R}^2$

which further gives rise to

$$\sum_{i=1}^{k} \frac{\lambda_i \bar{y}_i^2}{\lambda_{k+1}} + \sum_{j=k+1}^{n} \bar{y}_j^2 \leq \sum_{i=1}^{k} \left[(\bar{y}_i - \bar{\alpha}_i)^2 + \frac{\bar{\alpha}_i^2 \lambda_i}{\lambda_{k+1} - \lambda_i}\right] + \sum_{j=k+1}^{n} \bar{y}_j^2. \tag{16}$$

Since $\bar{y} \in \mathcal{B}(\bar{\alpha}, r^2(\bar{\alpha}))$, we have

$$\sum_{i=1}^{k} \left[(\bar{y}_i - \bar{\alpha}_i)^2\right] + \sum_{j=k+1}^{n} \bar{y}_j^2 \leq \frac{\gamma}{\lambda_{k+1}} - \sum_{i=1}^{k} \frac{\bar{\alpha}_i^2 \lambda_i}{\lambda_{k+1} - \lambda_i}. \tag{17}$$

Combining inequalities (16) and (17) yields

$$\sum_{i=1}^{k} \lambda_i \bar{y}_i^2 + \sum_{j=k+1}^{n} \lambda_{k+1} \bar{y}_j^2 \leq \gamma,$$

which implies $\bar{y} \in \mathcal{E}(\Lambda, \mathbf{0}, \gamma)$. We finally conclude that $\bigcup_{\alpha \in E_k} \mathcal{B}(\alpha, r^2) \subseteq \mathcal{E}(\Lambda, \mathbf{0}, \gamma)$.

□

Theorem 2.1 actually provides a parameterized representation of the ellipsoid $\mathcal{E}(\Lambda, \mathbf{0}, \gamma)$ by infinite number of balls, when Λ is of k-degree freedom. Figure 2 illustrates this decomposition scheme for an ellipsoid in $\mathbb{R}^2$ with 1-degree freedom. It is obvious that the center of the balls is along the longer radius of the ellipsoid and the union of infinite such balls is nothing but the ellipsoid itself.

We now proceed to extend the decomposition scheme in Theorem 2.1 to the ellipsoid $\mathcal{E}(Q, h, \rho)$ in constraint (8), when Q is of k-degree freedom. For any $\rho \geq 0$, we first decompose ellipsoid $\mathcal{E}(\Lambda^Q, \mathbf{0}, \rho)$ as follows,

$$\mathcal{E}(\Lambda^Q, \mathbf{0}, \rho) = \bigcup_{\alpha \in E_k^Q} \mathcal{B}(\alpha, r^2(\alpha)),$$

where

$$E_k^Q \triangleq \left\{ u \in \mathbb{R}^n \mid \sum_{i=1}^{k} \kappa_i u_i^2 \leq \rho,\ u_j = 0, j = k+1, \cdots, n \right\}, \tag{18}$$

$$r^2(\alpha) \triangleq (\rho - \sum_{i=1}^{k} (\iota_i \alpha_i^2))/\lambda_{k+1}^Q, \tag{19}$$

$$\kappa_i \triangleq \frac{\lambda_i^Q (\lambda_{k+1}^Q)^2}{(\lambda_{k+1}^Q - \lambda_i^Q)^2}, \text{ for } i = 1, \cdots, k, \tag{20}$$

$$\iota_i \triangleq \frac{(\lambda_i^Q \lambda_{k+1}^Q)}{\lambda_{k+1}^Q - \lambda_i^Q}, \text{ for } i = 1, \cdots, k. \tag{21}$$

As the shape and size of ellipsoid $\mathcal{E}(\Lambda^Q, 0, \rho)$ are coordinate independent, the affine transformation $x = \Gamma' y + h$ maps $y \in \mathcal{E}(\Lambda, 0, \rho)$ to $x \in \mathcal{E}(Q, h, \rho)$. Thus, the constraint (8) in problem $(\mathcal{P})$ can be expressed as follows according to Theorem 2.1,

$$x \in \mathcal{E}(Q, h, \rho) = \bigcup_{\alpha \in E_k^Q} \mathcal{B}(\Gamma' \alpha + h, r^2(\alpha)),$$

where E_k^Q and $r^2(\alpha)$ are defined by (18) and (19), respectively. Furthermore, problem $(\mathcal{P}_1)$ can be reformulated as,

$$(\mathcal{P}_2): \min_{\alpha, \rho} \ \frac{1}{2}\rho + C,$$

$$\text{Subject to: } x \in \bigcup_{\alpha \in E_k^Q} \{x \mid \lambda_{k+1} \|x - \Gamma' \alpha - h\|_2^2 + \sum_{i=1}^{k} \iota_i \alpha_i^2 \leq \rho\}, \tag{22}$$

$$x \in \Delta(s). \tag{23}$$

The formulation of problem $(\mathcal{P}_2)$ is still hard to solve, as the constraint in (22) involves an infinite number of balls. As we pointed out before, minimizing $f(x)$ under constraint (1) is equivalent to finding the minimum ellipsoid that touches the set $\Delta(s)$. We further recognize here that, among infinite balls which together make up the minimum ellipsoid, one specific ball offers the tangent point to achieve this task. Please refer to Figure 1 again, in which the red ball serves exactly this purpose. Based on this argument, we now construct the following problem to identify this particular ball in order to solve problem $(\mathcal{P}_2)$, albeit indirectly,

$$(\hat{\mathcal{P}}): \quad \min_{\rho, \beta} \ \frac{1}{2}\rho + C,$$

$$\text{Subject to: } \lambda_{k+1} \text{dis}(\beta) + \sum_{i=1}^{k} \iota_i \beta_i^2 \leq \rho, \tag{24}$$

$$\sum_{i=1}^{k} \kappa_i \beta_i^2 \leq \rho, \tag{25}$$

where decision variables are $\beta \in \mathbb{R}^k$, $\rho \in \mathbb{R}$ and the distance function $\mathrm{dis}(\cdot) : \mathbb{R}^k \to \mathbb{R}_+$ defined as

$$\mathrm{dis}(\beta) \triangleq \min_{x \in \mathbb{R}^n} \left\{ \|x - H\beta - h\|_2^2 \mid x \in \Delta(s) \right\}, \tag{26}$$

with $H \in \mathbb{R}^{n \times k}$ being formed by taking the first k columns of Γ'. Let $H_i \in \mathbb{R}^{1\times k}$ be the i-th row of matrix H and h_i be the i-th element of h.

Define $b(\beta) = (b_1(\beta), b_2(\beta), \cdots, b_n(\beta))'$, where

$$b_i(\beta) \triangleq |H_i\beta + h_i|, \text{ for } i = 1, \cdots, n. \tag{27}$$

Definition 2.1. We define an index set $\mathcal{I}(\beta) \subset \{1, 2, \cdots, n\}$ as the set that includes indices of the first $n - s$ smallest elements in $b(\beta)$ and the complementary set of $\mathcal{I}(\beta)$ as $\bar{\mathcal{I}}(\beta) = \{1, \cdots, n\} \setminus \mathcal{I}(\beta)$.

Note that for any fixed β, $\mathcal{I}(\beta)$ may not be unique. When we have multiple candidates of $b_i(\beta)$ to be chosen as the $(n-s)$th smallest element in $b(\beta)$ or the last element in $\mathcal{I}(\beta)$, we can take an arbitrary choice and this does not affect our discussion.

Theorem 2.2. *If solution-pair $(\hat{\beta}, \hat{\rho})$ solves problem $(\hat{\mathcal{P}})$, then solution-triple $(\hat{\alpha}, \hat{\rho}, \hat{x})$ solves problem $(\mathcal{P}_2)$, where $\hat{\alpha} = ((\hat{\beta})', \mathbf{0}'_{n-k})'$, $\hat{x} = (\hat{x}_1, \cdots, \hat{x}_n)'$ with*

$$\hat{x}_i = \begin{cases} H_i\hat{\beta} + h_i & i \in \bar{\mathcal{I}}(\hat{\beta}), \\ 0 & i \in \mathcal{I}(\hat{\beta}), \end{cases} \tag{28}$$

and index sets $\mathcal{I}(\hat{\beta})$ and $\bar{\mathcal{I}}(\hat{\beta})$ are defined in Definition 2.1.

Proof. Substituting $(\hat{\alpha}, \hat{\rho}, \hat{x})$ into (22) and (23) confirms the feasibility of $(\hat{\alpha}, \hat{\rho}, \hat{x})$ in problem $(\mathcal{P}_2)$, thus giving rise to $v(\mathcal{P}_2) \le v(\hat{\mathcal{P}})$. Now we assume that solution-triple $(\bar{\alpha}, \bar{\rho}, \bar{x})$ solves problem $(\mathcal{P}_2)$ with $v(\mathcal{P}_2) = \frac{1}{2}\bar{\rho} + C < v(\hat{\mathcal{P}}) = \frac{1}{2}\hat{\rho} + C$. Since $(\bar{\alpha}, \bar{\rho}, \bar{x})$ is the solution of problem $(\mathcal{P}_2)$, $(\bar{\alpha}, \bar{\rho})$ satisfies constraint (25) in problem $(\hat{\mathcal{P}})$. On the other hand, note that the following inequality always holds,

$$\lambda_{k+1}(\min_x \|x - H\bar{\beta} - h\|_2^2) + \sum_{i=1}^k \iota_i \bar{\beta}_i^2 \le \lambda_{k+1}\|\bar{x} - \Gamma'\bar{\alpha} - h\|_2^2 + \sum_{i=1}^k \iota_i \bar{\alpha}_i^2 \le \bar{\rho}, \tag{29}$$

where $\bar{\alpha}$ and $\bar{\beta}$ satisfy $\bar{\alpha} = (\bar{\beta}', 0'_{n-k})'$. Inequality (29) implies that $(\bar{\alpha}, \bar{\rho})$ satisfies constraint (24). From our assumption that $\frac{1}{2}\bar{\rho} + C < \frac{1}{2}\hat{\rho} + C$, we find a better solution $(\bar{\alpha}, \bar{\rho}, \bar{x})$ for problem $(\hat{\mathcal{P}})$, which contradicts the optimality of solution $(\hat{\alpha}, \hat{\rho})$ for problem $(\hat{\mathcal{P}})$. Note that once optimal $\hat{\beta}$ is fixed, the optimal $\hat{x}$ can be found by minimizing (26), i.e., we have

$$\mathrm{dis}(\hat{\beta}) = \min_{x \in \Delta(s)} \sum_{i=1}^n (x_i - H_i\hat{\beta} - h_i)^2. \tag{30}$$

Since $x \in \Delta(s)$, we choose $\hat{x}_i = H_i\hat{\beta} + h_i$, for $i \in \bar{\mathcal{I}}$ and $x_i = 0$ for $i \in \mathcal{I}$ which minimizes $\mathrm{dis}(\beta)$ in (30). ∎

Theorem 2.2 reveals an equivalence between problems $(\hat{\mathcal{P}})$ and $(\mathcal{P}_2)$. While we will focus on problem $(\hat{\mathcal{P}})$ in the following, the key issue we are facing is how to handle function $\mathrm{dis}(\beta)$ defined in (26). Clearly, function $\mathrm{dis}(\beta)$ measures the distance between the affine space, $\{y \in \mathbb{R}^n \mid y = H\beta + h\}$ and the set $\Delta(s)$. We will carry out detailed discussion on how to identify function $\mathrm{dis}(\beta)$ in Section 3.

3 Distance Evaluation

Given $h \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n\times k}$ with $rank(H) = k$, the distance function, $\mathrm{dis}(\cdot)$, is defined in (26). In this section, we focus on distance function $\mathrm{dis}(\beta)$, which plays a key role in solving problem $(\hat{\mathcal{P}})$. In particular, we show that $\mathrm{dis}(\beta)$ is a piece-wise continuous convex quadratic function and all the coefficients can be found explicitly by implementing our proposed algorithm. Geometrically, $\mathrm{dis}(\beta)$ identifies the minimum distance between an affine space and the feasible set $\Delta(s)$,

$$\mathrm{dis}(\beta) = \min_{x\in\mathbb{R}^n} \left\{\|y - x\|^2 \mid y \in \mathcal{Y}_k(H, h), x \in \Delta(s)\right\},$$

where $\mathcal{Y}_k(H, h) = \left\{y \in \mathbb{R}^n \mid y = H\beta + h, \beta \in \mathbb{R}^k\right\}$. Although the number of feasible subspaces in $\Delta(s)$ is of a combinatorial nature, the function $\mathrm{dis}(\beta)$ can be still characterized efficiently.

We now borrow some concepts from the discrete geometry [1] by considering the following hyperplane arrangements generated by the following hyperlanes in $\mathbb{R}^k$,

$$p^1_{i,j} \triangleq \{\beta \in \mathbb{R}^k \mid (H_i + H_j)\beta + (h_i + h_j) = 0\}, \tag{31}$$

$$p^2_{i,j} \triangleq \{\beta \in \mathbb{R}^k \mid (H_i - H_j)\beta + (h_i - h_j) = 0\}, \tag{32}$$

for $(i, j) \in \mathbb{I} \triangleq \{(i, j)|i = 1, \cdots, n - 1, j = i + 1, \cdots, n\}$. Note that the total number of such hyperplanes is $n(n - 1)$. A cell E of the hyperplane arrangements corresponding to $p^1_{i,j}$ and $p^2_{i,j}$ is a k-dimensional polyhedral set formed by the half spaces induced by hyperlanes $p^1_{i,j}$ and $p^2_{i,j}$, $(i, j) \in \mathbb{I}$.

We characterize the positive and negative half spaces of $p^1_{i,j}$ and $p^2_{i,j}$, respectively, by

$$w^1_{i,j} = \begin{cases} + & \text{if } (H_i + H_j)\beta + (h_i + h_j) \geq 0 \\ - & \text{if } (H_i + H_j)\beta + (h_i + h_j) < 0 \end{cases}, \text{ for } (i, j) \in \mathbb{I}, \tag{33}$$

$$w^2_{i,j} = \begin{cases} + & \text{if } (H_i - H_j)\beta + (h_i - h_j) \geq 0 \\ - & \text{if } (H_i - H_j)\beta + (h_i - h_j) < 0 \end{cases}, \text{ for } (i, j) \in \mathbb{I}. \tag{34}$$

Thus, any cell can be characterized by an $(n \times n)$ up-triangular sign matrix w,

$$\mathrm{sign}(E) = w = \begin{pmatrix} 0 & w_{1,2} & w_{1,3} & \cdots & w_{1,n} \\ & 0 & w_{2,3} & \cdots & w_{2,n} \\ & & \cdots & \cdots & \cdots \\ & & & & w_{n-1,n} \\ & & & & 0 \end{pmatrix}, \tag{35}$$

where $w_{i,j}$ is specified by

$$w_{i,j} = (w^1_{i,j} \circ w^2_{i,j}) \tag{36}$$

and the operator "$\circ$" is defined such that $(+ \circ +) = +$, $(+ \circ -) = -$, $(- \circ +) = -$ and $(- \circ -) = +$.

Lemma 3.1. *In each cell E, induced by hyperplane arrangements $p^1_{i,j}$ and $p^2_{i,j}$, $(i,j) \in \mathbb{I}$, the order of functions $\{b_i(\beta)\}_{i=1}^n$ is invariant within the cell, i.e., for a permutation of index set $\{1, 2, \cdots, n\}$, $\{i_1, i_2, \cdots, i_n\}$, the following holds true when β varies within cell E,*

$$b_{i_1}(\beta) \le b_{i_2}(\beta) \cdots \le b_{i_n}(\beta).$$

Proof. Clearly, the order of $\{b_i(\beta)\}_{i=1}^n$ is determined by comparing whether $b_i(\beta) - b_j(\beta) \ge 0$ or not, for all pair $(i,j) \in \mathbb{I}$. Since $b_i(\beta) \ge 0$ for all i, checking whether $b_i(\beta) - b_j(\beta) \ge 0$ or not is equivalent to checking whether the difference $b_i(\beta)^2 - b_j(\beta)^2$ is nonnegative or not. Note that, for any $(i,j) \in \mathbb{I}$, we have

$$(b_i(\beta))^2 - (b_j(\beta))^2 = \big((H_i + H_j)\beta + (h_i + h_j)\big)\big((H_i - H_j)\beta + (h_i - h_j)\big), \tag{37}$$

which further implies

$$(b_i(\beta))^2 - (b_j(\beta))^2 \begin{cases} \ge 0 & \text{if } (w^1_{i,j} = +, w^2_{i,j} = +) \text{ or } (w^1_{i,j} = -, w^2_{i,j} = -), \\ \le 0 & \text{if } (w^1_{i,j} = -, w^2_{i,j} = +) \text{ or } (w^1_{i,j} = +, w^2_{i,j} = -). \end{cases}$$

As any point β in cell E possesses the same sign vector $\text{sign}(E)$, thus the order of $\{b_i(\beta)\}|_{i=1}^n$ is invariant within each cell E. □

When $k = 0$, the affine space $\mathcal{Y}_k(H,h)$ degenerates to a singleton h, i.e., $\mathcal{Y}_0 = \{y \in R^n \mid y = h\}$, and $\{b\}|_{i=1}^n = \{|h_i|\}|_{i=1}^n$. As the index set $\mathcal{I}(\beta)$, in such a case, includes the indices corresponding to the first $n - s$ smallest elements of $\{|h|_i\}|_{i=1}^n$, the distance function $\text{dis}(\beta)$ becomes a constant which can be explicitly expressed.

Lemma 3.2. *When $k = 0$, the projection of h on $\Delta(s)$ is $x^* = \{x^*_i\}|_{i=1}^n = \arg\min_{x \in \Delta(s)} \|h - x\|_2^2$ with*

$$x^*_i = \begin{cases} h_i & i \in \bar{\mathcal{I}}(\beta) \\ 0 & i \in \mathcal{I}(\beta) \end{cases}$$

and $dis(\beta) = \sum_{i \in \mathcal{I}(\beta)} (h_i)^2$.

Note that the projection point may not be unique. We continue to prove that, when $k \ge 1$, the distance function is a piece-wise quadratic function. Since each cell of a hyperplane arrangement is a polyhedra, we use a unified expression $\Psi_t \beta \le \eta_t$, where $\Psi_t \in \mathbb{R}^{m \times k}$ and $\eta_t \in \mathbb{R}^m$, for cell t. While m is always bounded as $m \le n(n-1)$, it is bounded from above more tightly by the number of hyperplanes which are active for the concerned cell. We will describe in details our algorithm in Section 3.1 and Section 3.2 to search for all cells.

Theorem 3.1. *The distance function dis*(β) *is a piece-wise continuous quadratic function, with the following quadratic form with respect to* β *for each cell indexed by* $t \in \{1, \cdots, N\}$, *i.e.,*

$$dis(\beta) = \beta' D_t \beta + d_t \beta + c_t, \ \forall \beta \text{ satisfying } \Psi_t \beta \leq \eta_t, \tag{38}$$

where $D_t \in \mathbb{S}^k_{++}, d_t \in \mathbb{R}^k$ *and* $c_t \in \mathbb{R}$. *Furthermore, the total number of cells,* N, *is bounded from above by* $\mathcal{O}\left((n^2 - n)^k\right)$.

Proof. For any fixed $\tilde{\beta}$, applying Lemma 3.2 gives rise to

$$\begin{aligned} \text{dis}(\tilde{\beta}) &= \min_{x \in \Delta(s)} \|H\tilde{\beta} + h - x\|_2^2 = \min_{x \in \Delta(s)} \sum_{i=1}^{n} (H_i' \tilde{\beta} + h_i - x_i)^2 \\ &= \sum_{i \in \mathcal{I}(\tilde{\beta})} b_i(\tilde{\beta})^2 = \tilde{\beta}' D \tilde{\beta} + d\tilde{\beta} + c, \end{aligned} \tag{39}$$

where

$$D = \sum_{i \in \mathcal{I}(\tilde{\beta})} H_i' H_i, \quad d = 2 \sum_{i \in \mathcal{I}(\tilde{\beta})} h_i H_i, \quad c = \sum_{i \in \mathcal{I}(\tilde{\beta})} h_i^2. \tag{40}$$

While the sets $\mathcal{I}(\beta)$ and $\bar{\mathcal{I}}(\beta)$ are completely determined by the order of $\{b_i(\beta)\}_{i=1}^n$, the order of functions $\{b_i(\beta)\}_{i=1}^n$ is invariant within each cell E induced by hyperplane arrangements in (31) and (32) according to Lemma 3.1. It has been known that the upper bound on the number of cells of the hyperplane arrangement generated by (31) and (32) is in order of $\mathcal{O}\left((n^2 - n)^k\right)$ (see [1]).

Now we prove the continuity of function dis(β). Since dis(β) is a continuous function in the interior of each cell E, we only have to check the boundary between cells. Without loss of generality, we consider two neighboring cells, E_1 and E_2, which are separated by hyperplane $p^1_{i^*,j^*}$. There exist two different cases: i) If E_1 and E_2 define a same index set $\mathcal{I}(\beta)$, then dis(β) is the same for both E_1 and E_2, which implies the continuity of dis(β). ii) Assume that cells of E_1 and E_2 define two different index sets $\mathcal{I}_1(\beta)$ and $\mathcal{I}_2(\beta)$. Since hyperplane $p^1_{i^*,j^*}$ separates these two cells, we know that $b_{i^*}(\beta)$ and $b_{j^*}(\beta)$ change order from the proof of Lemma 3.1. More specifically, the index sets $\mathcal{I}_1(\beta)$ and $\mathcal{I}_2(\beta)$ are different in two indices, i^* and j^*,

$$\begin{aligned} \mathcal{I}_1(\beta) &= \{i^*\} \cup (\mathcal{I}_1(\beta) \cap \mathcal{I}_2(\beta)), \\ \mathcal{I}_2(\beta) &= \{j^*\} \cup (\mathcal{I}_1(\beta) \cap \mathcal{I}_2(\beta)). \end{aligned}$$

Thus, we have

$$\begin{aligned} \text{dis}(\beta) &= \sum_{i \in \mathcal{I}_1(\beta) \cap \mathcal{I}_2(\beta)} b_i(\beta)^2 + b_{i^*}^2(\beta), \ \beta \text{ in cell } E_1, \\ \text{dis}(\beta) &= \sum_{i \in \mathcal{I}_1(\beta) \cap \mathcal{I}_2(\beta)} b_i(\beta)^2 + b_{j^*}^2(\beta), \ \beta \text{ in cell } E_2 \end{aligned}$$

Clearly, on the hyperplane $p^1_{i^*,j^*}$, which is the boundary between cells E_1 and E_2, we have $b^2_{i^*}(\beta) = b^2_{j^*}(\beta)$ from the definition (31) and relationship (37), which further implies that $\text{dis}(\beta)$ is continuous on the boundary $p^1_{i^*,j^*}$. □

From Theorem 3.1, we know that the distance function $\text{dis}(\beta)$ is a piece-wise quadratic function defined on the cells of hyperplane arrangements, which takes the form in (38) and the total number of the pieces, N, is bounded by $\mathcal{O}((n^2-n)^k)$. Although problem $(\hat{\mathcal{P}})$ is not convex, it can still be solved by evaluating N subproblems $\hat{\mathcal{P}}_t$, $t = 1, \cdots, N$, separately, as follows,

$$(\hat{\mathcal{P}}_t): \quad \min \ \frac{1}{2}\rho + C, \tag{41}$$

$$\text{Subject to: } \lambda^Q_{k+1}(\beta' D_t \beta + d_t \beta + c_t) + \sum_{i=1}^{k} \iota_i \beta_i^2 \leq \rho,$$

$$\sum_{i=1}^{k} \kappa_i \beta_i^2 \leq \rho, \tag{42}$$

$$\Psi_t \beta \leq \eta_t.$$

Once we solve all these subproblems, we can identify the optimal solution (β^*, ρ^*) by solving a particular sub-problem $(\hat{\mathcal{P}}_{t^*})$, where

$$t^* = \arg \min_{t=1,\cdots,N} v(\hat{\mathcal{P}}_t).$$

Note that when β^* is fixed, we simply use formulation (28) to identify the optimal solution of problem $(\mathcal{P})$ with optimal objective $v(\mathcal{P}) = \frac{1}{2}\rho^* + C$.

Remark 3.1. From Corollary 3.1 in Section 3, when $k = 1$, the total number of the quadratic pieces of $\text{dis}(\beta)$ has a bound, $N \leq n(n-1)$. Furthermore, the sub-problem $(\hat{\mathcal{P}}_t)$ can be simplified to the following problem,

$$(\hat{\mathcal{P}}_t): \quad \min_{\beta,\rho} \ \tau = \frac{1}{2}\rho + C,$$

$$\text{Subject to: } \lambda^Q_2(D_t\beta^2 + d_t\beta + c_t) + \iota_1\beta^2 \leq \rho, \tag{43}$$

$$\kappa_1\beta^2 \leq \rho, \tag{44}$$

$$I_t \leq \beta \leq I_{t+1}.$$

Compared with the general case with $k > 1$, a more efficient algorithm is devised in Section 3.1 to identify the distance function $\text{dis}(\beta)$ for $k = 1$.

Remark 3.2. Initial bounds of β are critical for identifying the distance function $\text{dis}(\beta)$, as they affect significantly the speed of the search procedure described in Section 3. Generally speaking, when $Q \succ 0$, such bounds on β can be obtained from the following observation. In problem $(\hat{\mathcal{P}})$, the constraint in (42), $\sum_{i=1}^{k} \kappa_i\beta_i \leq \rho$, implies $\sum_{i=1}^{k} \kappa_i\beta_i \leq 2\overline{v(\mathcal{P})} + 2C$, where $\overline{v(\mathcal{P})}$ is an upper bound of problem $(\mathcal{P})$. In other words, we are only interested in identifying the distance function dis(β) on some bounded domain of β. From an algorithmic point of view, we prefer box-type of bound on β. Thus, we may assume β is confined in a box ,

$$\beta \in [\omega^l, \omega^u] \triangleq \left\{ \beta \in \mathbb{R}^k \mid \omega_i^l \le \beta_i \le \omega_i^u,\ i = 1, \cdots, k \right\},$$

where $\omega^l = \{\omega_i^l\}|_{i=1}^n$ and $\omega^y = \{\omega_i^b\}|_{i=1}^n$. An upper bound $\overline{v(\mathcal{P})}$ of problem ($\mathcal{P}$) can be easily found by some heuristics, e.g., from the objective value of the incumbent (the best feasible solution obtained).

Although Theorem 3.1 shows that the function dis(β) has at most $\mathcal{O}((n^2 - n)^k)$ pieces, in real application, the total number of the pieces is far less than this upper bound, especially, when we add box bound on β (see Remark 3.2). In the following subsections, we focus on developing an algorithm to identify the function dis(β) in a bounded domain of β, i.e., to identify the coefficients, $\{D_t, d_t, c_t\}$, $t = 1, \cdots, N$, for $\beta \in [w^l, w^u]$. We separate our discussion for cases of $k = 1$ and $k > 1$.

3.1 Identification of dis(β) for $k = 1$

When $k = 1$, both H and h are vectors in $\mathbb{R}^k$.

Corollary 3.1. *When $k = 1$, the distance function dis(β) consists of at most $N \le n(n-1)$ pieces of quadratic functions.*

Proof. The proof of the corollary follows Theorem 3.1. However, when $k = 1$, cells of the hyperplane arrangement degenerate to intervals on a real line. An upper bound of the total number of cells, N, can be calculated. Clearly, the set $\mathcal{I}(\beta)$ changes only when some $b_i(\beta)$ intersects with $b_j(\beta)$, with $i \in \mathcal{I}(\beta)$ and $j \in \bar{\mathcal{I}}(\beta)$. That is to say, $N \le S_n$, where S_n is the total number of intersection points between the functions $b_i(\beta)$ and $b_j(\beta)$ for $i \ne j$, and $i, j = 1, \cdots, n$. The number S_n can be computed in a recursive way. When $n = 2$, it holds that $S_2 = 2$, and, when $n > 3$, the recursion $S_n = S_{n-1} + 2(n-1)$ holds. Solving such a recursion yields $S_n = (n-1)n$. Thus, theoretically, we can divide the interval $[-\infty, \infty]$ into at most $N \le n(n-1)$ consecutive intervals $[I_j, I_{j+1}]$, for $j = 1, \cdots, N$. $\square$

From Corollary 3.1, we know that $N \le n(n-1)$, where notation "O" is dropped. However, it is still expensive and unnecessary to compute all N quadratic functions directly.

To identify function dis(β) with $\beta \in [\omega^l, \omega^u]$, we partition the interval $[\omega^l, \omega^u]$ into several sub-intervals, where functions $\{b_i(\beta)\}|_{i=1}^n$ are linear in each of there sub-intervals. Note that such a partition always exists. Since $b_i(\beta)$ is a constant when $H_i = 0$, we assume that $H_i \ne 0$ for all $i = 1, \cdots, n$. Function $b_i(\beta)$ achieves its minimum point at $-h_i/H_i$ with $b_i(\beta) = 0$, for $i = 1, \cdots, n$. If $-h_i/H_i \notin [\omega^l, \omega^u]$, then $b_i(\beta)$ is linear in $[\omega^l, \omega^u]$. Without loss of generality, we assume that points $\{-h_i/H_i\}|_{i=1}^n$ are arranged in an ascending order and are all in interval $[\omega^l, \omega^u]$,

$$\omega^l < -\frac{h_1}{H_1} \le -\frac{h_2}{H_2} \le \cdots \le -\frac{h_n}{H_n} < \omega^u.$$

Clearly, in each interval of $[\omega^l, -\frac{h_1}{H_1}], \cdots, [-\frac{h_n}{H_n}, \omega^u]$, function $b_i(\beta)$ is linear with respect to β. (see Figure 3).

Now we concentrate on identifying $\text{dis}(\beta)$ in each sub-interval within $\beta \in [\beta^l, \beta^u]$. Our main scheme is to sequentially check the intersection point between $b_i(\beta)$, $i \in \mathcal{I}(\beta)$ and $b_j(\beta)$, $j \in \bar{\mathcal{I}}(\beta)$ from β^l to β^u. Once an intersection point is identified, the index set $\mathcal{I}(\beta)$ is modified accordingly. More specifically, we use the following Table, $\mathbb{T}$, to store the data.

T	$\bar{I}_1$	$\bar{I}_2$	$\cdots$	$\bar{I}_s$
I_1	$T_{1,1}$	$T_{1,2}$	$\cdots$	$T_{1,s}$
I_2	$T_{2,1}$	$T_{2,2}$	$\cdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
I_{n-s}	$T_{n-s,1}$	$\cdots$	$\cdots$	$T_{n-s,s}$

In table T, the first column and the first row are corresponding to the index sets $\mathcal{I}_\beta$ and $\bar{\mathcal{I}}_\beta$, respectively, i.e., $I_i \in \mathcal{I}(\beta)$, $i = 1, \cdots, n-s$ and $\bar{I}_i \in \bar{\mathcal{I}}(\beta)$ for $i = 1, \cdots, s$. Element $\mathbb{T}(i,j)$ in the table stores the intersection point of $b_{I_i}(\beta)$ and $b_{I_j}(\beta)$. Since any two linear functions intersect at most once, each time we only need to modify one column and one row, which leads to a linear time operation $\mathcal{O}(n)$. We present formally such an algorithm in Algorithm 1 and use the following Example 3.1 to illustrate the procedure.

Algorithm 1 Procedure for identifying $\text{dis}(\beta)$ when $k = 1$

Input: Interval $[\omega^l, \omega^u]$, H and h

Output: All individual pieces of function $\text{dis}(\beta)$

(1) Let $\beta^l \leftarrow \omega^l$, $\beta^u \leftarrow \omega^u$, flag $\leftarrow 1$. Sort $\{b_i(\beta^l)\}|_{i=1}^n$ and construct $\mathcal{I}(\beta^l)$ and $\bar{\mathcal{I}}(\beta^l)$.

(2) Initialize Table T by filling first column and first row with index set $\mathcal{I}(\beta^l)$ and $\bar{\mathcal{I}}(\beta^l)$. For $i = 1, \cdots, n-s$ and $j = 1, \cdots, s$, set

$$T_{i,j} = \begin{cases} \frac{h_{I_i} - h_{I_j}}{H_{I_j} - H_{I_i}} & \text{if } \beta^l \le \frac{h_{I_i} - h_{I_j}}{H_{I_j} - H_{I_i}} \le \beta^u, \\ +\infty & \text{otherwise}, \end{cases} \tag{45}$$

while flag $= 1$ **do**
  Construct $\mathcal{I}(\beta^l)$ and identify $\text{dis}(\beta)$ as given in (39) and (40). Output $\text{dis}(\beta)$ with $[\beta^l, \beta^u]$.
  if $T_{i,j} \neq \infty$ for all $i = 1, \cdots, n-s$, $j = 1, \cdots, s$, **then**
    Find the minimum value T_{i^*,j^*} in Table T. $\beta^l \leftarrow \beta^u$, $\beta^u \leftarrow T_{i^*,j^*}$ and $T_{i^*,j^*} \leftarrow \infty$. Exchange index I_{j^*} and I_{i^*} and update j^*-th column and i^*-th row by using (45).
    if $\beta^u = \omega^u$, **then**
      flag $\leftarrow 0$
    end if
  else
    flag $\leftarrow 0$
  end if
end while

Example 3.1. We consider an example with $n = 6$, $k = 1$, $s = 3$ and

$$H = \begin{pmatrix} -0.5 & 0.5 & 1 & -0.75 & -2 & -4 \end{pmatrix}',$$
$$h = \begin{pmatrix} 1 & 2.5 & 4 & 3.4 & 3.5 & 5.2 \end{pmatrix}'.$$

We want to identify function $\text{dis}(\beta)$ with $\beta \in [-1, 1.3]$.

For this example, functions $\{b_i(\beta)\}_{i=1}^6$ are specified as follows (also see Figure 3),

$$\begin{array}{ll} b_1(\beta) = |-0.5\beta + 1|, & b_2(\beta) = |0.5\beta + 2.5|, \\ b_3(\beta) = |\beta + 4|, & b_4(\beta) = |-0.75\beta + 3.4|, \\ b_5(\beta) = |-2\beta + 3.5|, & b_6(\beta) = |-4\beta + 5.2|. \end{array}$$

Interval $[-5, 5]$ can be decomposed as follows,

$$[-5, 5] = \cup[-5, -4] \cup [-4, 1.3] \cup [1.3, 1.75] \cup [1.75, 2] \cup [2, 4.53] \cup [4.53, 5],$$

such that in each of the sub-intervals, all $b_i(\beta)$, $i = 1, \cdots, 6$, are linear.

We demonstrate our algorithm, in particular, for the interval of $[-1, 1.3]$ in which

$$\begin{array}{ll} b_1(\beta) = -0.5\beta + 1, & b_2(\beta) = 0.5\beta + 2.5, \\ b_3(\beta) = \beta + 4, & b_4(\beta) = -0.75\beta + 3.4, \\ b_5(\beta) = -2\beta + 3.5, & b_6(\beta) = -4\beta + 5.2. \end{array}$$

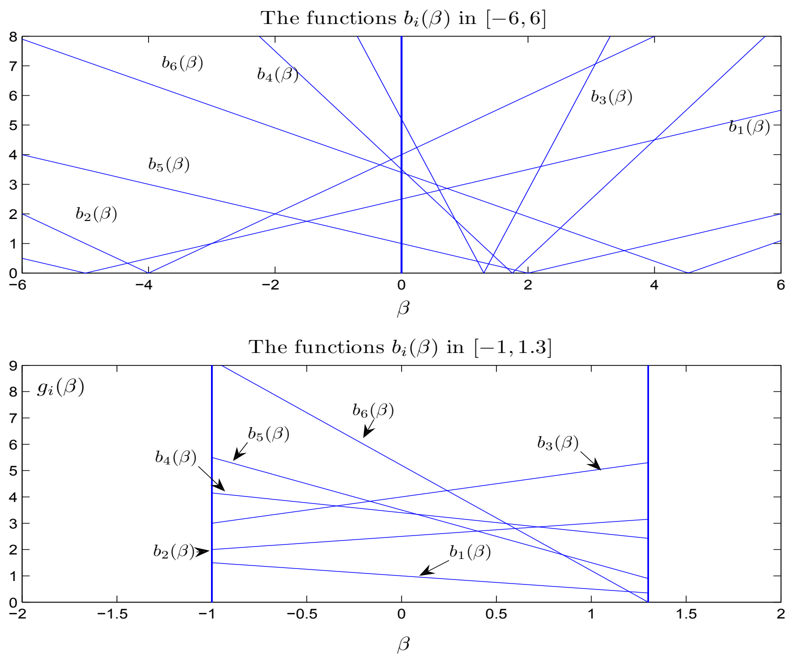

Figure 3: Functions $b_i(\beta)$ in Example 3.1

We now use Algorithm 1 to identify function $\text{dis}(\beta)$ in the interval of $[-1, 1.3]$.

- In the first step, we compute $b_i(-1)$ for $i = 1, \cdots, 6$, initialize the index sets as $\mathcal{I}(-1) = \{1, 2, 3\}$ and $\bar{\mathcal{I}}(-1) = \{4, 5, 6\}$ according to the values of $\{b_i(-1)\}|_{i=1}^{6}$, and construct Table 1, where the element in the position corresponding to $b_i \in \mathcal{I}$ and $b_j \in \bar{\mathcal{I}}$ represents their intersection. We find the minimum value -0.342 in Table 1, which indicates that

Table 1: Table of Step 1 in Example 3.1

	6	4	5
1	1.200	$+\infty$	$+\infty$
2	0.600	0.720	0.400
3	0.240	-0.343	-0.167

$b_4(\beta)$ and $b_3(\beta)$ intersecting at $\beta = -0.342$. We can thus conclude that, in the sub-interval of $\beta \in [-1, -0.342]$, the index set $\mathcal{I}(\beta) = \{1, 2, 3\}$ remains unchanged. We further identify the first piece of function $\text{dis}(\beta)$ for $\beta \in [-1, -0.342]$ as

$$\text{dis}(\beta) = 1.5\beta^2 + 9.5\beta + 23.25, \quad \beta \in [-1, -0.342],$$

where the coefficients are computed according to (39) with index set $\mathcal{I}([-1, -0.342]) = \{1, 2, 3\}$. Since -0.342 is the intersection point of $b_3(\beta)$ and $b_4(\beta)$, we exchange the positions of b_3 by b_4 in Table 1, which leads to Table 2.

Table 2: Table of Step 2 in Example 3.1

	6	3	5
1	1.200	$+\infty$	$+\infty$
2	0.600	$+\infty$	0.400
4	0.554	$+\infty$	0.008

- The minimum value in Table 2 is 0.08 and we derive the second piece of $\text{dis}(\beta)$ as

$$\text{dis}(\beta) = 1.062\beta^2 - 36\beta + 18.51, \quad \beta \in [-0.342, 0.080], \tag{46}$$

where all coefficients are computed according to index set $\mathcal{I}([-0.342, 0.08]) = \{1, 2, 4\}$. Updating row 3 and column 3 in Table 2 yields Table 3.

Table 3: Table of Step 3 in Example 3.1

	6	3	4
1	1.200	$+\infty$	$+\infty$
2	0.600	$+\infty$	0.720
5	0.850	$+\infty$	$+\infty$

- The minimum value of Table 3 is 0.60 and we derive the third piece of function $\text{dis}(\beta)$ as

$$\text{dis}(\beta) = 4.5\beta^2 - 12.5\beta + 19.5, \quad \beta \in [0.08, 0.60], \tag{47}$$

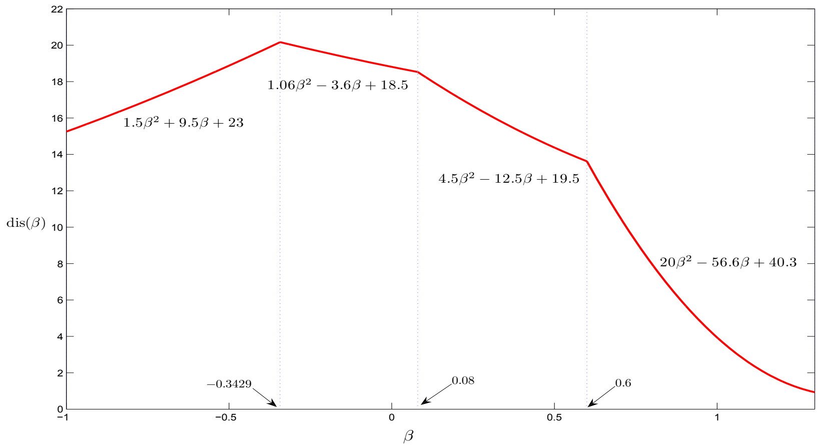

Figure 4: Functions $\{b_i(\beta)\}|_{i=1}^6$

where all coefficients are computed according to index set $\mathcal{I}([0.08, 0.6]) = \{1, 2, 5\}$. Updating row 2 and column 1 in Table 3 yields Table 4.

Table 4: Table of Step 4 in Example 3.1

	2	3	4
1	$+\infty$	$+\infty$	$+\infty$
6	$+\infty$	$+\infty$	$+\infty$
5	$+\infty$	$+\infty$	$+\infty$

- The minimum value of Table 4 is 1.3 and we get the fourth piece of function dis(β) as

$$\text{dis}(\beta) = 20.25\beta^2 - 56.6\beta + 40.29, \quad \beta \in [0.60, 1.3]. \tag{48}$$

As all elements in Table 4 are ∞, we complete the characterization of the distance function of dis(β) in $[-1, 1.3]$ (See Figure 4).

3.2 Identification of dis(β) for $k > 1$

We know from Theorem 3.1 that the distance function dis(β) can be constructed according to the cells of hyperplane arrangement. Enumerating the cells of the hyperplane arrangement has been investigated in the literature. For example, the authors in [1] and [10] proposed a cell enumeration method by reverse searching method. Such a method consumes $\mathcal{O}((n^2 - n)^k C_{lp})$ time to enumerate all the cells, where C_{lp} is the time for a linear programming. Note that the cells are searched in the whole $\mathbb{R}^k$ space when adopting the reverse searching method. In our study, we are only interested in enumerating all the cells within a bounded region $\beta \in [\omega^l, \omega^u]$, which is also ready to be solved by the algorithm in [10].

As we have illustrated in Lemma 3.1, the order of all $\{b_i(\beta)\}|_{i=1}^n$ is fixed in each cell induced by hyperplanes $p^1_{i,j}$ and $p^2_{i,j}$, for $i, j \in \mathbb{I}$. Such cells are characterized by the sign matrix sign(E) in (35). Now we specify a mapping from a sign matrix to an order of functions $\{b_i(\beta)\}|_{i=1}^n$. Given a sign matrix sign(E) in (35), we first construct matrix $\Omega(E)$ by letting $\Omega_{ii}(E) = 0$, copying the upper-triangle of sign(E) into its upper-triangle and inserting the opposite of the upper-triangle of sign(E) into its lower-triangle.

From (37), (36) and (33), we can conclude that $b_i(\beta)$ is the $(t+1)$-th smallest element among $\{b_i(\beta)\}_{i=1}^n$ for β in cell E if there are totally t of "+" in i-th row of $\Omega(E)$, i.e., we have

$$\underbrace{(+,-,+,\cdots,-)}_{\text{There are } t \text{ "+"s.}} \longleftrightarrow \ b_i(\beta) \text{ is } (t+1)\text{-th smallest elemetnt.}$$

Example 3.2. We illustrate how to identify function dis(β) with $k > 1$ by considering the following example of identifying function dis(β) with $k = 2$, $n = 4$, $s = 2$ and parameters set as follows,

$$H = \begin{pmatrix} 4 & 2 \\ 5 & 1 \\ 1 & -1 \\ 2 & -0.5 \end{pmatrix}, \quad h = \begin{pmatrix} -2 \\ -6 \\ -1 \\ 2 \end{pmatrix}.$$

Assume that $\beta = (\beta_1, \beta_2)'$ is confined in the box of $-1 \le \beta_1 \le 4$ and $-1.5 \le \beta_2 \le 1$. According to (31) and (32), we introduce the following hyperplanes,

$$\begin{array}{ll}
p^1_{1,2}: \ z^1_{1,2}(\beta) = -\beta_1 + \beta_2 + 4 = 0, & p^2_{1,2}: \ z^2_{1,2}(\beta) = 9\beta_1 + 3\beta_2 - 8 = 0, \\
p^1_{1,3}: \ z^1_{1,3}(\beta) = 3\beta_1 + 3\beta_2 - 1 = 0, & p^2_{1,3}: \ z^2_{1,3}(\beta) = 5\beta_1 + \beta_2 - 3 = 0, \\
p^1_{1,4}: \ z^1_{1,4}(\beta) = 2\beta_1 + 2.5\beta_2 - 4 = 0, & p^2_{1,4}: \ z^2_{1,4}(\beta) = 6\beta_1 + 1.5\beta_2 = 0, \\
p^1_{2,3}: \ z^1_{2,3}(\beta) = 4\beta_1 + 2\beta_2 - 5 = 0, & p^2_{2,3}: \ z^2_{2,3}(\beta) = 6\beta_1 - 7 = 0, \\
p^1_{2,4}: \ z^1_{2,4}(\beta) = 3\beta_1 + 1.5\beta_2 - 8 = 0, & p^2_{2,4}: \ z^2_{2,4}(\beta) = 7\beta_1 + 0.5\beta_2 - 4 = 0, \\
p^1_{3,4}: \ z^1_{3,4}(\beta) = -\beta_1 - 0.5\beta_2 - 3 = 0, & p^2_{3,4}: \ z^2_{3,4}(\beta) = 3\beta_1 - 1.5\beta_2 + 1 = 0.
\end{array}$$

It can be verified that the region specified by $-1 \le \beta_1 \le 4$ and $-1.5 \le \beta_2 \le 1$ is on one side of the following hyperplane,

$$z^2_{1,3}(\beta) > 0, \ z^2_{1,4}(\beta) > 0, \ z^2_{2,4}(\beta) > 0, \ z^1_{3,4}(\beta) < 0, \ z^2_{3,4}(\beta) > 0.$$

Thus, for any β in this region, it holds true that $\omega^2_{1,3} = +$, $\omega^2_{1,4} = +$, $\omega^2_{2,4} = +$, $\omega^1_{3,4} = -$, $\omega^2_{3,4} = +$. We can enumerate the cells of the arrangements generated by these hyperplanes in the box region, $\beta_1 \in [1, 4]$ and $\beta_2 \in [-1.5, 1]$. Implementing the algorithm of cell enumeration [10] generates 15 cells. While the sign vectors of these 15 cells are listed in Table 5, all the hyperplanes and their arrangement are illustrated in Figure 5, in which the arrow indicates the positive side of the hyperplane. We can then figure out the order of the $\{b_i(\beta)\}_{i=1}^4$ in each cell. For example, let us consider $cell_2$ in Table 5 with

$$\text{sign}(cell_2) = \begin{pmatrix} & (-\circ+) & (+\circ+) & (+\circ+) \\ & & (+\circ+) & (+\circ+) \\ & & & (-\circ+) \\ & & & \end{pmatrix}.$$

Then we construct matrix $\Omega(cell_2)$ from the sign matrix sign$(cell_2)$ as

$$\Omega(cell_2) = \begin{pmatrix} 0 & - & + & + \\ + & 0 & + & + \\ - & - & 0 & - \\ - & - & + & 0 \end{pmatrix}.$$

which yields an order of $b_i(\beta)$, i.e., $b_3(\beta) < b_4(\beta) < b_1(\beta) < b_2(\beta)$. Once the order of $b_i(\beta)$ is achieved, the distance function can be expressed by applying (39) and (40),

$$\text{dis}(\beta) = 5\beta_1^2 + 1.25\beta_2^2 - 8\beta_1\beta_2 + 6\beta_1 + 5.$$

All the other pieces of the distance function can be derived in a similar fashion.

Table 5: The cells of hyperplanes in Example 3.2

No	$(w_{1,2}, w_{1,3}, w_{1,4}), (w_{2,3}, w_{2,4}), (w_{3,4})$
1	$(++,++,++),(++,++),(-+)$
2	$(-+,++,++),(++,++),(-+)$
3	$(-+,++,-+),(++,++),(-+)$
4	$(-+,++,-+),(++,+-),(-+)$
5	$(++,++,-+),(++,+-),(-+)$
6	$(++,++,-+),(-+,+-),(-+)$
7	$(++,+-,-+),(-+,+-),(-+)$
8	$(+-,+-,-+),(-+,+-),(-+)$
9	$(+-,+-,-+),(--,+-),(-+)$
10	$(+-,--,-+),(--,+-),(-+)$
11	$(+-,--,-+),(--,++),(-+)$
12	$(+-,--,-+),(-+,++),(-+)$
13	$(+-,--,-+),(+-,++),(-+)$
14	$(+-,--,++),(+-,++),(-+)$
15	$(+-,--,++),(++,++),(-+)$

4 Illustrative Example

We demonstrate in this section a complete implementation of our solution scheme developed in this paper via an illustrative example.

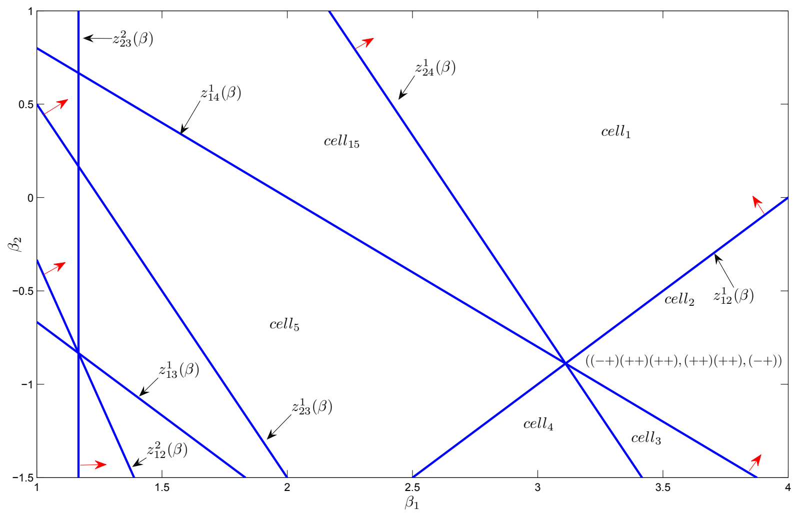

Figure 5: The cells of hyperplane arrangement in Example 3.2

Example 4.1. Let us consider an example of problem $(\mathcal{P})$ with $n = 6$, $s = 2$ and

$$Q = \begin{pmatrix} 27.171 & -5.738 & 2.479 & -2.768 & 4.931 & 1.725 \\ -5.738 & 18.358 & 5.030 & -5.615 & 10.004 & 3.500 \\ 2.479 & 5.030 & 27.827 & 2.426 & -4.322 & -1.512 \\ -2.768 & -5.615 & 2.426 & 27.292 & 4.825 & 1.688 \\ 4.931 & 10.004 & -4.322 & 4.825 & 21.404 & -3.007 \\ 1.725 & 3.500 & -1.512 & 1.688 & -3.007 & 28.948 \end{pmatrix},$$
$$q = \begin{pmatrix} 37.745 & -26.329 & -80.284 & 34.905 & 7.296 & -51.002 \end{pmatrix}'.$$

It can be verified that Q is of a $k = 1$- freedom with one eigenvalue of $\lambda_1 = 1$ and five eigenvalues of $\lambda_2 = 30$. For this simple example, adoption of an enumeration method identifies its optimal solution of $x^* = (0, 0, 2.989, 0, 0, 1.918)'$ and the corresponding optimal value of $v(\mathcal{P}) = -168.91$.

We compute H and h defined (26) as

$$H = (\ 0.312,\ 0.634,\ -0.274,\ 0.306,\ -0.544,\ -0.190\ )',$$
$$h = (\ -11.369,\ -19.636,\ 11.539,\ -11.057,\ 17.384,\ 7.867)'.$$

Using the parameters defined in Section 2.2, we have $\kappa_1 = 1.0701$, $\iota_1 = 1.0344$ and $C = -749.43$, which are defined in (20), (21) and (5), respectively. Since $k = 1$, we identify the distance function $\mathrm{dis}(\beta)$ as follows in the interval $[0, 37]$ by using the method discussed in

Section 3.1,

$$\mathrm{dis}(\beta) = \begin{cases} 9.063\beta^2 - 695.19\beta + 13396.41, & \mathcal{I}(\beta) = \{2,5\}, \quad \beta \in [0, 21.59], \\ 15.71\beta^2 - 1073.53\beta + 18467.86, & \mathcal{I}(\beta) = \{2,3\}, \quad \beta \in [21.59, 25.26], \\ 18.86\beta^2 - 1252.13\beta + 20969.04, & \mathcal{I}(\beta) = \{3,5\}, \quad \beta \in [25.26, 25.91], \\ 24.83\beta^2 - 1606.95\beta + 26156.91, & \mathcal{I}(\beta) = \{3,1\}, \quad \beta \in [25.91, 28.75], \\ 26.66\beta^2 - 1730.09\beta + 28178.20, & \mathcal{I}(\beta) = \{3,6\}, \quad \beta \in [28.75, 33.37], \\ 15.71\beta^2 - 1073.53\beta + 18467.85, & \mathcal{I}(\beta) = \{2,3\}, \quad \beta \in [33.37, 35.34], \\ 9.06\beta^2 - 695.19\beta + 13396.41, & \mathcal{I}(\beta) = \{2,5\}, \quad \beta \in [35.34, 36.60]. \end{cases} \tag{49}$$

Note that function $\mathrm{dis}(\beta)$ consists of total $N = 7$ pieces of convex quadratic functions. We can now explicitly write out the left hand of constraint (43) for each of sub-problem $(\hat{\mathcal{P}}_t)$, $t = 1, \cdots, 7$,

$$\lambda_2 \mathrm{dis}(\beta_1) + \iota_1 \beta_1^2 = \begin{cases} g_1(\beta) = 10.10\beta^2 - 695.19\beta + 13396.41 & \beta \in [0, 21.59], \\ g_2(\beta) = 16.74\beta^2 - 1073.53\beta + 18467.86 & \beta \in [21.59, 25.26], \\ g_3(\beta) = 19.89\beta^2 - 1252.13\beta + 20969.04 & \beta \in [25.26, 25.91], \\ g_4(\beta) = 25.86\beta^2 - 1606.95\beta + 26156.91 & \beta \in [25.91, 28.75], \\ g_5(\beta) = 27.70\beta^2 - 1730.09\beta + 28178.20 & \beta \in [28.75, 33.37], \\ g_6(\beta) = 16.74\beta^2 - 1073.53\beta + 18467.85 & \beta \in [33.37, 35.34], \\ g_7(\beta) = 10.10\beta^2 - 695.19\beta + 13396.41 & \beta \in [35.34, 36.60]. \end{cases}$$

On the other hand, constraint (44) keeps the same form, $\bar{g}(\beta) = 1.07\beta^2 \le \rho$. The sub-problem $(\hat{\mathcal{P}}_t)$, $t = 1, \cdots, 7$, becomes

$$(\hat{\mathcal{P}}_t): \min_{\beta,\rho} \left\{ \frac{1}{2}\rho + C \mid g_t(\beta) \le \rho, \quad \bar{g}(\beta) \le \rho \right\}.$$

We plot all the constraints in Figure 6. Solving all these sub-problems, we find that the optimal value is achieved in interval $[25.91, 28.75]$ with index set $\mathcal{I}(\beta) = \{3, 1\}$, the optimal ρ^* equal to 1161.053 and the optimal value $v(\mathcal{P}) = \frac{1}{2}\rho^* + C = -168.91$. From Theorem 2.2, we further get optimal solution $x_3^* = 2.989$, $x_6^* = 1.918$, $x_1^* = 0$, $x_2^* = 0$, $x_4^* = 0$, $x_5^* = 0$.

5 Conclusions

We have investigated a class of CCQO problems with their coefficient matrices being of a k-degree freedom. We have demonstrated that we can decompose such a CCQO problem into several quadratic convex subproblems, where the total number of the subproblems is determined by the number of the cells generated by some $\mathbb{R}^k$ dimensional hyperplanes within a bounded region. It is interesting to note that such a number is bounded by $\mathcal{O}(n^{2k})$. In particular, for the case with $k = 1$, we have developed an efficient method to identify all the subproblems. For cases with $k \ge 2$, by modifying the reverse searching method proposed in [1], we have also developed an efficient algorithm to identify the subproblems. Once subproblems are known, we can solve each subproblem individually and find the solution of the original CCQO problem.

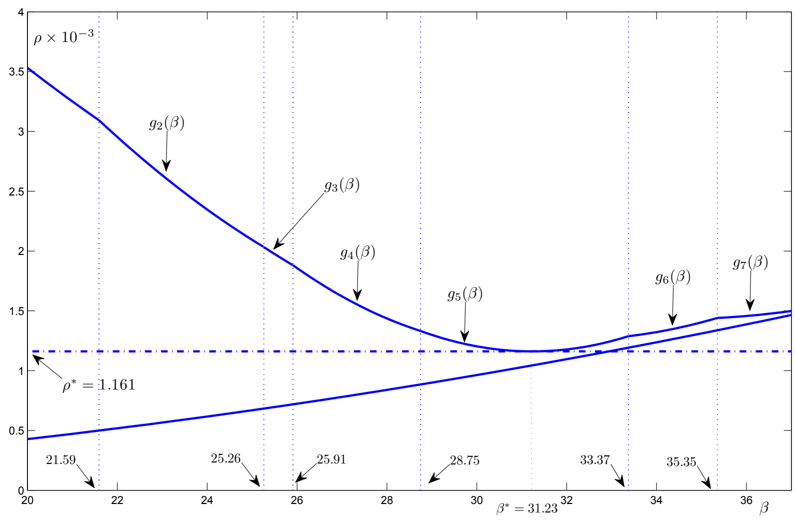

Figure 6: Functions $\{g_i(\beta)\}|_{i=1}^{6}$ and $\bar{g}(\beta)$ of Example 4.1

References

[1] D. Avis and K. Fukuda. Reverse search for enumeration. *Discrete Applied Mathematics*, 65:21–46, 1996.

[2] D. Bertsimas and R. Shioda. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications*, 2007.

[3] D. Bienstock. A computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74:121–140, 1996.

[4] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete freguency information. *IEEE Transcation on Information Theory*, 52(2):489–509, 2006.

[5] A. Das and D. Kempe. Algorithms for subset selection in linear regression. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 45–54, 2008.

[6] D. L. Donoho. Compressed sensing. *IEEE Transaction Information Theory*, 52(4):1289, 2006.

[7] M. R. Garey and D. S. Johnson. *Computer and Intractability, A Guide to The Thoery of NP-Completeness*. W. H. Freeman Co, Francisco, 1979.

[8] T. M. Liebling E. Steiner K. Allemand, K. Fukuda. A polynomial case of unconstrained zero-one quadratic optimization. *Mathematical Programming*, 91:49–52, 2001.

[9] D.X. Shawa, S. Liub, and L. Kopmanb. Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optimization Methods and Software*, 23:411–420, 2008.

[10] N. Sleumer. Output-sensitive cell enumeration in hyperplane arrangements. *Nordic journal of computing*, 6:137–161, 1999.

[11] W.J. Welch. Algorithm complexity: three np-hard problems in computational statistics. *Journal of Statistical Computation and Simulation*, 15:17–25, 1982.

[12] J. Xie, S. He, and S. Zhang. Randomized portfolio selection with constraints. *Pacific Journal of Optimization*, 4:89–112, 2008.