Vector Space Optimization: EE 8950

Murti V. Salapaka
Electrical Engineering Department
Univeristy of Minnesota
murtis@umn.edu

We will begin by giving the following definitions:

Definition 1. [Group] *A group is a set G with a binary operation $(.) : G \times G \to G$ defined which has the following properties.*

1. $(a.b).c = a.(b.c)$; *associativity property.*

2. *There exists an element e in G such that $a.e = e.a = a$ for all a in G. e is called the identity.*

3. *For every a in G there exists an element a^{-1} in G such that $a.a^{-1} = a^{-1}.a = e$. a^{-1} is called the inverse of a.*

Definition 2. [Subgroup] *If H is a subset of a group G the H is a subgroup if H is a group with the binary operation inherited from G.*

Lemma 1. *H is a subgroup of the group G if the identity element e is in H, a belongs to H implies a^{-1} is in H and a and b belong to H implies $a.b$ belongs to H.*

Lemma 2. *A group G has a unique identity element. Also, every element in G has a unique inverse.*

Definition 3. [Abelian group] *A group G is an abelian group if for any two elements in G, $a.b = b.a$.*

Definition 4. [Homomorphism] *Let G and H be two groups. $\phi : G \to H$ is a homomorphism between the two groups if $\phi(a.b) = \phi(a).\phi(b)$, for all a, b in G.*

Lemma 3. *A homomorphism $\phi : G \to H$ sends identity of G to the identity of H and sends inverses to inverses.*

Definition 5. [Isomorphism] *An isomorphism is a homomorphism which is one to one and onto.*

Definition 6. [Fields] *A field K is a set that has the operations of addition $(+) : K \times K \to K$ and multiplication $(.) : K \times K \to K$ defined such that*

1. *multiplication distributes over addition*

$$a.(b + c) = a.b + a.c,$$

2. *K is an abelian group under addition with identity written as 0 for addition.*

3. *$K \backslash \{0\}$ is an abelian group under multiplication with identity being 1.*

Lemma 4. *If in a field K elements $a \neq 0$ and $b \neq 0$ then $ab \neq 0$.*

Vector Space

Definition 7. *A set V with two operations addition $(+) : V \times V \to V$ and scalar multiplication $(.) : V \times K \to V$, where K is a field defined is a vector space over the field K if*

1. *V is an abelian group under addition.*

2. *multiplication distributes over addition*

$$\alpha.(b + c) = \alpha.a + \alpha.b, \text{ for all } \alpha \text{ in } K, \text{ for all } a, b \text{ in } V.$$

The elements of the field K are often called as scalars. The vector space is called a real vector space if the field $K = R$ and the vector space is called a complex vector space if the field $K = C$.

Definition 8. [Algebra] *V a vector space is an algebra if it has an operation vector multiplication $(\cdot) : V \times V \to V$ defined such that this operation distributes over vector addition.*

Definition 9. [Units] *If A is an algebra then x in A is an unit if there exists some y in A such that $x \cdot y = y \cdot x = 1$.*

Lemma 5. *If A is an algebra with an associative vector multiplication and U is the set of units in A then U is a group under vector multiplication.*

From now on we will restrict the field to be either the set of real numbers R or the set of complex numbers C. Thus when we say K we mean either R or C.

Normed Vector Space

Definition 10. *A normed linear space is a vector space X with a function $||\cdot|| : X \to R$ defined such that*

1. $||x|| \geq 0$ *and* $||x|| = 0$ *if and only if* $x = 0$.

2. $||\alpha x|| = |\alpha|\ ||x||$ *for any scalar* α *and vector* x *in* X.

3. $||x + y|| \leq ||x|| + ||y||$.

Convergence in a Normed Vector Space

Definition 11. [Convergence] *Let V be vector space with a norm $\|.\|$ defined. Suppose v_n is a sequence in V, then v_n converges to some $v \in V$, if $\|v_n - v\| \to 0$ as $n \to \infty$.*

Cauchy Sequence

Definition 12. *A sequence v_n in a normed vector space V is said to be cauchy if given $\epsilon > 0$, there exists an integer N such that if $n, m \geq N$ then $\|v_n - v_m\| \leq \epsilon$.*

Lemma 6. *Every Convergent sequence is Cauchy.*

Complete Normed Space; Banach Space

Definition 13. *A normed vector space in which every Cauchy sequence is convergent, is called complete vector Space.*

Pre-Hilbert Space

Definition 14. *A vector space with inner product* $< .,. >$ *defined.*

Orthogonal Vectors

Definition 15. *In a Pre-Hilbert Space, two vectors x and y said to be orthogonal if* $< x, y >= 0$*, (x* $\perp$ *y). Moreover, if* $< x, x >= 1$ *and* $< y, y >= 1$*, then x and y are called orthonormal vectors.*

Cauchy-Schwartz Inequality

Lemma 7. *[Cauchy-Schwartz inequality] : For all* x, y *in a Pre-Hilbert space* $| < x, y > | \leq \|x\|\|y\|$. *Furthermore, the equality holds iff either* $y = 0$ *or* $x = \lambda y$ *where* λ *is a scalar.*

Continuity of Inner Product

Theorem 1. *Let* $x_n \in H$ *and* $y_n \in H$ *be sequences in a pre-Hilbert space* H *such that*

$$\|x_n - x\| \to 0 \textit{ as } n \to \infty$$

and

$$< h, y_n > \to < h, y > \textit{ for all } h \in H \textit{ and } \|y_n\| \le M \textit{ for all } n.$$

(Thus $x_n \to x$ *in norm topology and* $y_n \to y$ *in the weak-star topology.) Then*

$$< x_n, y_n > \to < x, y > .$$

Proof: Note that

$$\begin{aligned}
|<x_n, y_n> - <x, y>| &= |<x_n, y_n> - <x, y_n> + <x, y_n> - <x, y>| \\
&= |<x_n - x, y_n> + <x, y_n - y>| \\
&\leq |<x_n - x, y_n>| + |<x, y_n - y>| \\
&\leq |<x_n - x, y_n>| + |<x, y_n - y>| \\
&\leq \|x_n - x\| \|y_n\| + |<x, y_n - y>| \\
&\leq \|x_n - x\| M + |<x, y_n - y>|
\end{aligned}$$

Given $\epsilon > 0$ choose N such that $n > N$ implies that $\|x_n - x\| \leq \frac{\epsilon}{2M}$ and $|<x, y_n - y>| \leq \frac{\epsilon}{2}$. Thus given $\epsilon > 0$ there exists a N such that $n \geq N$ implies that

$$|<x_n, y_n> - <x, y>| \leq \epsilon.$$

∎

The following lemma is a special case of the above theorem:

Lemma 8. *In a Pre-Hilbert Space H, suppose v_n and w_n are sequences in H converging to v and w respectively $(\in H)$. Then, $< v_n, w_n > \;\to\; < v, w >$ as $n \to \infty$.*

Proof: Follows easily from the fact that if

$$\|w_n - w\| \to 0 \Rightarrow < h, w_n > \to < h, w >$$

and that $\|w_n\|$ is uniformly bounded. ∎

Hilbert Space

Definition 16. *Let H be a Pre-Hilbert Space with inner product* $< .,. >$ *defined. Then let* $\|x\|$ = $\sqrt{< x, x >}$ *defines the norm. If H together with the norm is Complete Space, then H is a Hilbert Space.*

Some Results

Lemma 9. *A set* $S \subset \{X, \|.\|\}$ *is closed iff it contains all its limit points.*

Theorem 2. *Every finite dimensional vector space is closed.*

Pre-Hilbert Space Theorem

Theorem 3. *Let H be a pre-Hilbert space and let M be a subspace of H. Then for $m_0 \in M$*

$$\|x - m_0\| \le \|x - m\| \text{ for all } m \in M \text{ if and only if } (x - m_0) \perp M.$$

Proof: ($\Leftarrow$) Suppose $m_0 \in M$ is such that $(x - m_0) \perp M$. Then for any $m \in M$ it follows that

$$\begin{aligned}
\|x - m\|^2 &= \|x - m + m_0 - m_0\|^2 \\
&= \|(x - m_0) + (m - m_0)\|^2 \\
&= \|(x - m_0)\|^2 + \|(m - m_0)\|^2 \\
&\ge \|x - m_0\|^2
\end{aligned}$$

The third equality holds because because $(x - m_0) \in M^{\perp}$ and $(m - m_0) \in M$.

($\Rightarrow$) Let $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$. Suppose $x - m_0 \notin M^{\perp}$. Then there exists $m_1 \in M$ with $\|m_1\| = 1$ and

$$< x - m_0, m_1 >= \delta > 0.$$

Let $\tilde{m} := m_0 + \delta m_1$. It follows that

$$\begin{aligned} < x - \tilde{m}, x - \tilde{m} > &= < x - m_0 - \delta m_1, x - m_0 - \delta m_1 > \\ &= \|x - m_0\|^2 + \delta^2 < m_1, m_1 > -2\delta < x - m_0, m_1 > \\ &= \|x - m_0\|^2 + \delta^2 - 2\delta\delta \\ &< \|x - m_0\|^2 \end{aligned}$$

Note that $\tilde{m} \in M$. This contradicts the hypothesis and thus $\delta = 0$. $\blacksquare$

Classical Projection Theorem

Theorem 4. *Let H be a Hilbert space and M a closed subspace of H. Consider the following problem*

$$\mu = \inf\{\|x - m\|,\ m \in M\},$$

where $x \in H$. Then, there exists a unique vector $m_0 \in M$ such that $\|x - m_0\| = \mu$, that is,

$$m_0 = arg\{\inf_{m \in M} \|x - m\|\}$$

Furthermore, a necessary and sufficient condition for m_0 being the unique minimizing vector is

$$(x - m_0) \perp M.$$

Proof: Note that only the existence of m_0 needs to be proven. The rest of the proof follows from Theorem 3.

If $x \in M$, then $m_0 = x$ and theorem is proven. Suppose $x \notin M$. Then for any $n \in N$. there exists $m_n \in M$ such that $\mu \leq \|x - m_n\| \leq \mu + \frac{1}{n}$. Thus there exist a sequence $\{m_i\}_{i=0}^{\infty} \in M$ such that $\|x - m_n\|$ converges to μ as $n \to \infty$. From the parallelogram law, for any integer i and j,

$$\|(m_j - x) + (x - m_i)\|^2 + \|(m_j - x) - (x - m_i)\|^2 = \ 2\|m_j - x\|^2 + 2\|m_i - x\|^2.$$

This implies that

$$\|m_j - m_i\|^2 + \|m_j + m_i - 2x\|^2 = \ 2\|m_j - x\|^2 + 2\|m_i - x\|^2.$$

Thus

$$\|m_j - m_i\|^2 = 2\|m_j - x\|^2 + 2\|m_i - x\|^2 - 4\|\frac{m_j + m_i}{2} - x\|^2. \tag{1}$$

Note that $\frac{m_j+m_i}{2} \in M$, and therefore

$$\|\frac{m_j + m_i}{2} - x\| \geq \ \mu.$$

From (1) we have

$$\|m_j - m_i\|^2 \leq 2\|m_j - x\|^2 + 2\|m_i - x\|^2 - 4\mu^2.$$

Given any $\epsilon > 0$, let N be a large positive integer such that for all $n \geq N$, $\|x - m_n\|^2 \leq \mu^2 + \frac{\epsilon^2}{4}$. If $i, j > N$ then,

$$\|m_j - m_i\|^2 \leq 2\mu^2 + \frac{\epsilon^2}{2} + 2\mu^2 + \frac{\epsilon^2}{2} - 4\mu^2.$$

This implies

$$\|m_j - m_i\|^2 \ \leq \epsilon^2.$$

It follows that

$$\|m_j - m_i\| \;\; \leq \;\; \epsilon$$

Thus, m_n forms a Cauchy Sequence. And, since M is a closed subspace of Hilbert Space (which is complete), m_n is a converging sequence (due to completeness) with the limit point inside M (due to closedness). Thus there exists a $m_0 \in M$, such that

$$\|m_n - m_0\| \to 0 \text{ as } n \to \infty.$$

Since, $\|(x - m_n) - (x - m_0)\| = \|m_n - m_0\|$ we have

$$(x - m_n) \to (x - m_0) \text{ as } n \to \infty$$

From the continuity of norm, $\|(x - m_n)\|$ converges to $\|(x - m_0)\|$. Since, a converging sequence has a unique limit point, we have

$$\mu = \|x - m_0\|.$$

This proves the theorem $\blacksquare$

Direct Sum

Definition 17. *A vector space X is said to be the direct sum of two subspaces M and N if every vector $x \in X$ has a unique representation of the form $x = m + n$, where $m \in M$ and $n \in N$. The notation $X = M \oplus N$ is used to denote that X is a direct sum of M and N.*

Relationships between a space and its perp space

Theorem 5. *Let S and T be subsets of a Hilbert Space H, then*

1. $S^\perp$ *is a closed subspace.*

2. $S \subset S^{\perp\perp}$.

3. *If* $S \subset T$ *then* $T^\perp \subset S^\perp$.

4. $S^{\perp\perp\perp} = S^\perp$.

Proof: (1) Let p_n be a sequence in $S^\perp$ with $p_n \to p$. Let s be an element in S. As $p_n \in S^\perp$, it follows that

$$< s, p_n >= 0, \text{ for all } n.$$

Thus

$$\lim_{n\to\infty} < s, p_n >= 0.$$

Note that

$$\begin{aligned} | < s, p > | = | < s, p - p_n + p_n > | &= | < s, p_n > + < s, p - p_n > | \\ &\leq | < s, p_n > | + | < s, p_n - p| \\ &\leq | < s, p_n > | + \|s\| \, \|p - p_n\| \text{ for all } n. \end{aligned}$$

Taking limits we have that

$$| < s, p > | = \lim_{n\to\infty} | < s, p > | = \lim_{n\to\infty} | < s, p_n > | + \lim_{n\to\infty} \|s\| \, \|p - p_n\| = 0.$$

As s is an arbitrary element in S it follows that $p \in S^{\perp}$. This proves (1).

(2) Let s be any element in s. Then for all elements p in $S^{\perp} < s, p >= 0$. Therefore $s \in S^{\perp\perp}$. Thus $S \subset S^{\perp\perp}$. This proves (2).

(3) Let $t^{\perp}$ be an element in $T^{\perp}$. Then it follows that $< t, t^{\perp} >= 0$ for all $t \in T$. As $S \subset T$ it follows that $< t, t^{\perp} >=$ for all $t \in S$. Thus $t^{\perp} \in S^{\perp}$. As $t^{\perp}$ is an arbitrary element in $T^{\perp}$ it follows that $T^{\perp} \subset S^{\perp}$. This proves (3). (4) will be proven after establishing the next theorem. $\blacksquare$

Decomposition of a Hilbert space

Theorem 6. *If M is a closed linear subspace of a Hilbert space H, then*

$$H = M \oplus M^{\perp} \text{ and } M = M^{\perp\perp}.$$

Proof: Let h be an element in H. From the classical projection theorem (Theorem 4) it follows that there exists an element $h_m \in M$ such that

$$h_m = \arg\{\inf_{m \in M} \|x - m\|\}.$$

Furthermore, such an element is a unique element that satisfies $(h - h_m) \in M^{\perp}$. Let $h_n := h - h_m$. Then clearly $h = h_m + h_n$ with $h_m \in M$ and $h_n \in M^{\perp}$.

Suppose $h = h'_m + h'_n$ is another decomposition with $h'_m \in M$ and $h'_n \in M^{\perp}$.

Then it follows that $h = h_m + h_n = h'_m + h'_n$ and therefore

$$(h'_m - h_m) + (h'_n - h_n) = 0.$$

As, $(h'_m - h_m) \in M$ and $(h'_n - h_n) \in M^\perp$ it follows that

$$0 = \|(h'_m - h_m) + (h'_n - h_n)\|^2 = \|(h'_m - h_m)\|^2 + \|(h'_n - h_n)\|^2.$$

Thus $h'_m = h_m$ and $h'_n = h_n$. Thus the decomposition $h = h_m + h_n$ is unique. This proves $H = M \oplus M^\perp$.

In Theorem 5 we have established that $M \subset M^{\perp\perp}$. Suppose M is closed. Let $p \in M^{\perp\perp}$. From the decomposition result established earlier p can be decomposed as $p = p_m + p_n$ with $p_m \in M$ and $p_n \in M^\perp$. It follows that

$$p_n = p - p_m,$$

and

$$\|p_n\|^2 =< p_n, p_n >=< p - p_m, p_n >=< p, p_n > - < p_m, p_n >= 0.$$

This follows because $< p, p_n >= 0$ (as $p \in M^{\perp\perp}$ and $p_n \in M^{\perp}$) and $< p_m, p_n >= 0$ (as $p_n \in M^{\perp}$ and $p_m \in M$). Thus $p_n = 0$ and

$$p = p_m + p_n = p_m \in M.$$

Thus we have shown that any arbitrary element $p \in M^{\perp\perp}$ also belongs to M. Therefore $M^{\perp\perp} \subset M$. This proves the theorem. ■

Minimum Norm Vector of a Linear Variety (Restatement of the Projection Theorem)

Theorem 7. *Let V be a linear variety defined by*

$$V = x + M = \{x + m | m \in M\}$$

where M is a closed subspace of a Hilbert space H and $x \in H$. For the optimization problem

$$\mu = \inf_{v \in V} \|v\|.$$

there exists a minimizing element $v_0 \in M^{\perp}$. That is there exists $v_0 \in M^{\perp}$ such that

$$v_0 = arg\{\inf_{v \in V} \|v\|\}.$$

Proof: Note that

$$\begin{aligned} \mu &= \inf_{v \in V} \|v\| \\ &= \inf_{m \in M} \|x + m\| \\ &= \inf_{m \in M} \|x - m\| \end{aligned}$$

From the classical projection theorem there exists a minimizing element m_0 to the problem

$$\inf_{m \in M} \|x - m\|$$

such that $x - m_0 \in M^{\perp}$. The proof of the theorem follows by defining $v_0 = x - m_0$ and from

$$arg\{\inf_{v \in V} \|v\|\} = x - arg\{\inf_{m \in M} \|x - m\|\} = x - m_0 = v_0.$$

■

Invertibility of the Gram Matrix

Definition 18. [Gram Matrix] *Gram matrix of set of vectors* $y_1, y_2 \ldots y_n$ *is given by :*

$$G(y_1, y_2 \ldots y_n) = \begin{pmatrix} < y_1, y_1 > & < y_1, y_2 > & \cdots & < y_1, y_n > \\ < y_2, y_1 > & < y_2, y_2 > & \cdots & < y_2, y_n > \\ \vdots & \vdots & \ddots & \vdots \\ < y_n, y_1 > & < y_n, y_2 > & \cdots & < y_n, y_n > \end{pmatrix}.$$

Lemma 10. $G(y_1, y_2 \ldots y_n)$ *is invertible if and only if the vectors* $y_1, y_2 \ldots y_n$ *are linearly independent.*

Proof:

Minimum Norm Vector that Satisfies a Set of Linear Equalities

Theorem 8. *Let* H *be a Hilbert space and* $y_i, \;\; i = 1, \ldots, n$ *be a set of linearly independent vectors in* H. *Let*

$$\mu \quad = \qquad\qquad \inf \|x\|$$

subject to

$$\begin{aligned} < x, y_1 > &= c_1 \\ < x, y_2 > &= c_2 \\ &\vdots \\ < x, y_n > &= c_n. \end{aligned}$$

Then there exists x_0 *to the above problem that is*

$$x_0 = arg\{\inf\{\|x\| \mid < x, y_i >= c_i, i = 1 \ldots, n\}\}$$

where x_0 is a linear combination of $y_{i's}$:

$$x_0 = \sum_{i=0}^{n} \beta_i y_i$$

with β_i satisfy following normal

$$G^T(y_1, y_2 \dots y_n) \underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}}_{:=\beta} = \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}}_{:=c}.$$

Proof: Let

$$\begin{aligned} Y &= span\{y_1, \dots, y_n\} \text{ and} \\ S &= \{x \in H | < x, y_i >= c_i, i = 1, \dots, n\}. \end{aligned}$$

It is evident that

$$\begin{aligned} Y^{\perp} &= \{x \in H| < x, y_i >= 0, i = 1, \ldots, n.\} \text{ and therefore} \\ S &= \overline{x} + Y^{\perp} \end{aligned}$$

where $\overline{x}$ is any element that belongs to S.

It is evident that

$$\begin{aligned} \mu &= \inf\{\|x\| \mid < x, y_i >= c_i, \ i = 1, \ldots, n\} \\ &= \inf_{x \in S} \|x\| \end{aligned}$$

where S is the linear variety $S = \overline{x} + Y^{\perp}$ (note that $Y^{\perp}$ is a closed subspace). From Theorem 7 it follows that there exists a minimizing solution $x_0 \in (Y^{\perp})^{\perp}$. Y being a finite dimensional vector space is closed (see Theorem 2.) Thus from Theorem 6 it follows that $x_0 \in Y^{\perp\perp} = Y$. Thus x_0 is a linear combination of the vectors y_i with

$$x_0 = \sum_{i=1}^{n} \beta_i y_i.$$

Also, $x_0 \in S$ and therefore for all $j = 1, \ldots, n$

$$c_j =< x_0, y_j >=< \sum_{i=1}^{n} \beta_i y_i, y_j >= \sum_{i=1}^{n} \beta_i < y_i, y_j > .$$

This set of relations is equivalent to the matrix equation $G^T \beta = c$. This proves the theorem. ■

Hilbert Space of Random Variables

Let (Ω, F, P) be a probability space. Let H be the the set of random variables with finite variance. H is endowed with an inner product $< X, Y >= E(XY)$. H together with the inner product is a Hilbert Space. Let H^n denote random vectors with dimension n.

Note that we are not endowing any Hilbert Space structure to H^n. The only Hilbert space we will be interested in is the **scalar** Hilbert space H.

Estimation

Consider two random vectors $\beta \in H^n$ and $y \in H^m$. where

The problem is to estimate the unknown random vector β from the measurements y. What is desired is a function $\hat{\beta} = f(y)$ which provides the estimate of β. Note that y is a random vector and as such we desire a mapping f that maps the m dimensional random variable to a n dimensional random vector. Thus a mapping from H^m to H^n is sought. Note that once an appropriate mapping f is determined for a *particular* realization of the random vector $y = y^1$ we can obtain the estimate *for that realization* as $f(y^1)$.

It needs to be emphasized that a mapping $f : H^m \to H^n$ is sought as an estimate of β. If the function f is restricted to be a linear function then the goal is to obtain a *linear estimate* of β given y.

Furthermore the estimate is termed *unbiased* if the function f is restricted such that $E(\hat{\beta}) = (\beta)$.

The estimate f_0 is said to be a *minimum variance estimate* if in the allowable class of mappings S

$$f_0 = arg\left[\inf_{f\in S} E\{(\beta - f)^T(\beta - f)\}\right].$$

Minimum-variance Estimate of X based on Y.

Let X and Y be two random variables. The following problem is of interest:

$$\mu = \inf_f E\{(X - f)^2\}.$$

Theorem 9. *Let X and Y be two random variables. Then*

$$E(X|Y) = arg\left[\inf_f E\{(X - f(Y))^2\}\right].$$

Proof: The proof provided is not rigorous; however, it conveys the main idea. Let

$$\mu = \inf_{f(Y)} E\{(X - f(Y))^2\}.$$

Let $M = \{m = f(Y) : \ m$ has finite variance$\}$. Note that every element of S is a random variable $m : \Omega \to R$ obtained by $m = f \circ Y$ where $f : R \to R$ and $Y : \Omega \to R$. M is a subspace of the Hilbert space H. Thus we have

$$\mu = \inf_{m \in M} \|X - m\|_2^2.$$

From the classical projection theorem (see Theorem 4) any m_0 which satisfies $(X - m_0) \perp M$ is the minimizer. Note that

$$\begin{aligned}
< X - E(X|Y), f(Y) > &= E[(X - E(X|Y))f(Y)] \\
&= E[Xf(Y)] - E[E(X|Y)f(Y)] \\
&= E[Xf(Y)] - E[E(f(Y)X|Y)] \\
&= 0.
\end{aligned}$$

Thus, $E(X|Y)$ is the vector that is perpendicular to all other vectors in M.

Thus the theorem follows from the classical projection theorem. ■

Note that identities like

$$E[Xf(Y)] = E[E(f(Y)X|Y)]$$

can be proven by assuming the pdf's $p_{x,y}(x, y),\ p_{x|y}(x|y),\ p(y)$ and $p(x)$ to represent the joint pdf of the random variables X and Y, the conditional pdf of X given Y, the marginal pdf of Y and the marginal pdf of X respectively.

Using this notation we have

$$
\begin{aligned}
E[Xf(Y)] &= \int\int xf(y)p_{x,y}(x,y)dxdy \\
&= \int\int xf(y)p_{x|y}(x|y)p(y)dxdy \\
&= \int\left(\int xf(y)p_{x|y}(x|y)dx\right)p(y)dy \\
&= \int E(f(y)X|Y=y)p(y)dy \\
&= E[E(f(Y)X|Y)]
\end{aligned}
$$

Similarly other such identities can be proven.

In the case where X and Y are jointly Gaussian one can show that $E(X|Y)$ is a *linear* function of the Gaussian variable Y.

It was seen that

- The minimum-variance estimate of X based on Y is $E(X|Y)$.

- In the case when X and Y have a jointly Gaussian distribution $E(X|Y)$ is linear in Y given by $R_{xy}R_y^{-1}Y$.

Without further knowledge of the joint pdf of X and Y it is not possible to evaluate $E(X|Y)$. Thus typically the minimum variance estimate is difficult to obtain. However, it is relatively straightforward to obtain a minimum-variance *linear* estimate of X based on Y as is seen by the theorem below.

Review of Gaussian Variables

Definition 19. [Gaussian random vector] *The random n dimensional vector X is said to Gaussian (normal) if it has a pdf described by*

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n |R_x|}} exp\{-\frac{1}{2}[x - m_x]^T R_x^{-1}[x - m_x]\}.$$

Theorem 10. *The random vector X with pdf*

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n |R_x|}} exp\{-\frac{1}{2}[x - m_x]^T R_x^{-1}[x - m_x]\}.$$

has the mean m_x and the covariance R_x. Thus

$$E_X(X) = m_x \text{ and } E_X([x - m_x][x - m_x]^T) = R_x.$$

Characteristic Functions: Generating higher order moments

Definition 20. *If X is a n dimensional random vector then the characteristic function $\phi_X(\cdot)$ is defined as a scalar function as below*

$$\phi_X(\mu) := E_X[e^{j\mu^T x}] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{j\sum_{i=1}^{n} \mu_i x_i} f_X(x) dx_1 \ldots dx_n.$$

Theorem 11.

$$E[x_{i_1} x_{i_2} \ldots x_{i_m}] = \frac{1}{j^m} \left[\frac{\partial^m \phi_X(\mu)}{\partial \mu_{i_1} \partial \mu_{i_1} \ldots \partial \mu_{i_m}} \right] \Bigg|_{\mu=0}$$

Proof: Note that

$$\phi_X(\mu) = \int exp(j\mu^T x) f_X(x) dx.$$

Therefore it follows that

$$\begin{aligned}
\frac{\partial \phi_X(\mu)}{\partial x_{i_1}} &= \frac{\partial}{\partial x_{i_1}} \left(\int exp(j\mu^T x) f_X(x) dx] \right) \\
&= \int \frac{\partial}{\partial x_{i_1}} (exp(j \sum_{i=1}^{n} \mu_i x_i) f_X(x)) dx \\
&= \int j x_{i_1} exp(j \sum_{i=1}^{n} \mu_i x_i) f_X(x)) dx
\end{aligned}$$

Differentiating the above expression with respect to x_{i_2} we obtain

$$\begin{aligned}
\frac{\partial^2 \phi_X(\mu)}{\partial x_{i_1} \partial x_{i_2}} &= \frac{\partial}{\partial x_{i_2}} \left(\int j x_{i_1} exp(j\mu^T x) f_X(x) dx] \right) \\
&= \int j x_{i_1} \frac{\partial}{\partial x_{i_2}} (exp(\sum_{i=1}^{n} \mu_i x_i) f_X(x)) dx \\
&= \int j^2 x_{i_1} x_{i_2} exp(\sum_{i=1}^{n} j \mu_i x_i) f_X(x)) dx
\end{aligned}$$

Proceeding in a similar manner one can show that

$$\frac{\partial^m \phi_X(\mu)}{\partial \mu_{i_1} \partial \mu_{i_1} \dots \partial \mu_{i_m}} = j^m \int x_{i_1} x_{i_2} \dots x_{i_m} exp(\sum_{i=1}^{n} j \mu_i x_i) f_X(x)) dx$$

The theorem follows by evaluating above at $\mu = 0$.

$\blacksquare$

Characteristic Functions of a sum of two independent random vectors

Theorem 12. *Let $Z = X + Y$ where X, Y are two n dimensional random vectors that are independent. Then*

$$\phi_Z(\mu) = \phi_X(\mu)\phi_Y(\mu).$$

Proof:

Note that

$$\phi_Z(\mu) = E[e^{j\mu^T Z}] = E[e^{j\mu^T (X+Y)}] = E[e^{j\mu^T X} e^{j\mu^T Y}]$$

and as X and Y are independent it follows that

$$E[e^{j\mu^T X} e^{j\mu^T Y}] = E[e^{j\mu^T X}]E[e^{j\mu^T Y}] = \phi_X(\mu)\phi_Y(\mu).$$

This proves the theorem.

∎

Characteristic function of a Gaussian random vector

Theorem 13. *The characteristic function of a n dimensional random vector X that has mean μ and variance R with a pdf given by*

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n|R|}} exp\{-\frac{1}{2}[x-m]^T R^{-1}[x-m]\}$$

is

$$\phi_X(\mu) = e^{j\mu^T m - \frac{1}{2}\mu^T R^{-1}\mu}.$$

Linear Transformation of a Gaussian vector is Gaussian

Theorem 14. *Let X be a n dimensional random vector with mean m_x and covariance R_x. Let $A \in R^{m \times n}$. Then $Y = AX$ is a m dimensional random vector that is Gaussian with mean m_y and covariance R_y where*

$$m_y = Am_x \text{ and } R_y = AR_x R^T.$$

Proof: The characteristic function of Y is given by

$$\begin{aligned} \phi_Y(\mu) = E[e^{j\mu^T Y}] = E[e^{j\mu^T AX}] = E[e^{j(A^T\mu)^T X}] &= \phi_X(A^T\mu) \\ &= exp\{j\mu^T Am_x - \tfrac{1}{2}\mu AR_x A^T \mu\} \end{aligned}$$

which is the characteristic function of a Gaussian vector with mean Am_x and variance $AR_x A^T$.

This completes the proof. ■

Jointly Gaussian Variables

Definition 21. [Jointly Gaussian] *Suppose X and Y are two random vectors of dimension n and m respectively. Let $Z := \begin{bmatrix} X \\ Y \end{bmatrix}$ be a $n+m$ dimensional random vector formed by stacking X and Y. X and Y are said to jointly Gaussian if Z is a Gaussian random vector of dimension $n+m$.*

Theorem 15. *Let X and Y be jointly Gaussian n and m dimensional random vectors with means m_x and m_y respectively and covariances R_x and R_y respectively. Let $A \in R^{p\times n}$ and $B \in R^{p\times m}$. Then $Z = AX + BY$ is a p dimensional random vector is Gaussian with mean*

$$m_z = Am_x + Bm_y \text{ and } R_z = AR_xA^T + AR_{xy}B^T + BR_{yx}A^T + BR_yB^T.$$

Proof:

Let $U = \begin{bmatrix} X \\ Y \end{bmatrix}$. As X and Y are jointly Gaussian, it follows that U is Gaussian with mean $m_u = \begin{bmatrix} m_x \\ m_y \end{bmatrix}$ and covariance

$$R_u = \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}.$$

Note that $Z = AX + BY = \underbrace{\begin{bmatrix} A & B \end{bmatrix}}_{H} U$. Therefore from Theorem 14 it follows that Z has the mean given by $Hm_u = Am_x + Bm_y$ and covariance

$$\begin{aligned} HR_uH^T &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix} \begin{bmatrix} A^T \\ B^T \end{bmatrix} \\ &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} R_x A^T + R_{xy} B^T \\ R_{yx} A^T + R_y B^T \end{bmatrix} \\ &= AR_xA^T + AR_{xy}B^T + BR_{yx}A^T + BR_yB^T. \end{aligned}$$

This proves the theorem.

■

Marginals from jointly Gaussian pdf

Theorem 16. *Suppose Z is a random vector of dimension $n+m$ with a pdf given as*

$$\begin{aligned} p_Z(z) &= [(2\pi)^{(n+m)/2}\left|\begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}\right|^{1/2}]^{-1} \\ & .exp\{-\tfrac{1}{2}\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}^T \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}^{-1} \begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}\}. \end{aligned}$$

where Z is partitioned as $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ with X and Y a n and m dimensional random vectors respectively. Then

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n|R_x|}} exp\{-\frac{1}{2}[x-m_x]^T R_x^{-1}[x-m_x]\}$$

and

$$p_Y(y) = \frac{1}{\sqrt{(2\pi)^m |R_y|}} exp\{-\frac{1}{2}[y - m_y]^T R_y^{-1} [y - m_y]\}.$$

Proof: Note that $X = \underbrace{\begin{bmatrix} I & 0 \end{bmatrix}}_{A} Z$. Therefore it follows from Theorem 14 that the mean of X is given by

$$Am_z = m_x$$

and the covariance is given by

$$AR_zA^T = R_x.$$

A similar derivation can be done to obtain the pdf of the random variable Y.

■

Theorem 17. *If X and Y are independent Gaussian vectors then they are jointly Gaussian.*

Proof: Note that if X and Y are Gaussian random vectors with dimensions n and m with pdfs given by

$$p_X(x) = \frac{1}{\sqrt{(2\pi)^n |R_x|}} exp\{-\frac{1}{2}[x - m_x]^T R_x^{-1} [x - m_x]\}$$

and

$$p_Y(y) = \frac{1}{\sqrt{(2\pi)^n |R_y|}} exp\{-\frac{1}{2}[y - m_y]^T R_y^{-1} [y - m_y]\}$$

then the joint pdf

$$
\begin{aligned}
p_{X,Y}(x,y) &= p_X(x)p_Y(y) \\
&= \frac{1}{\sqrt{(2\pi)^n|R_x|}} exp\{-\tfrac{1}{2}[x-m_x]^T R_x^{-1}[x-m_x]\} \\
&\quad .\frac{1}{\sqrt{(2\pi)^m|R_y|}} exp\{-\tfrac{1}{2}[y-m_y]^T R_y^{-1}[y-m_y]\} \\
&= \frac{1}{\sqrt{(2\pi)^n|R_x|(2\pi)^m|R_y|}} \\
&\quad .exp\{-\tfrac{1}{2}[x-m_x]^T R_x^{-1}[x-m_x] - \tfrac{1}{2}[y-m_y]^T R_y^{-1}[y-m_y]\} \\
&= \frac{1}{\sqrt{(2\pi)^{n+m}|R_x||R_y|}} \\
&\quad .exp\{-\tfrac{1}{2}\begin{bmatrix} (x-m_x)^T & (y-m_y)^T \end{bmatrix} \underbrace{\begin{bmatrix} R_x^{-1} & 0 \\ 0 & R_y^{-1} \end{bmatrix}}_{R_z^{-1}} \begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}\} \\
&= \frac{1}{\sqrt{(2\pi)^{n+m}|R_z|}} \\
&\quad .exp\{-\tfrac{1}{2}\begin{bmatrix} (x-m_x)^T & (y-m_y)^T \end{bmatrix} R_z^{-1} \begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}\}
\end{aligned}
$$

which is a Gaussian distribution with covariance R_z and mean $m_z = \begin{bmatrix} m_x \\ m_y \end{bmatrix}$.

$\blacksquare$

Uncorrelated implies independence for jointly Gaussian vectors

Definition 22. *X and Y two random vectors of dimensions n and m are uncorrelated if $E[(X - m_x)(Y - m_y)^T] = 0$ where $m_x = E[X]$ and $m_y = E[Y]$.*

Theorem 18. *Suppose X and Y are jointly Gaussian random vectors such that X and Y are not correlated. Then X and Y are independent.*

Proof: If X and Y are not correlated then $R_{xy} = R_{yx} = 0$. Note that

$$R_z = \begin{bmatrix} R_x & 0 \\ 0 & R_y \end{bmatrix}.$$

and $|R_z| = |R_x||R_y|$. The Gaussian distribution of $Z := \begin{bmatrix} X \\ Y \end{bmatrix}$ is given by (with

$z := \begin{bmatrix} x \\ y \end{bmatrix})$

$$
\begin{aligned}
p_Z(z) &= [(2\pi)^{(n+m)/2}|R_z|^{1/2}]^{-1} exp\{-\tfrac{1}{2}[z - z_m]^T R_z^{-1}[z - z_m]\} \\
&= [(2\pi)^{(n)/2}|R_x|^{1/2}]^{-1}(2\pi)^{(m)/2}|R_y|^{1/2}]^{-1} \\
&\quad .exp\{-\tfrac{1}{2}[x - m_x]^T R_x^{-1}[x - m_x] - \tfrac{1}{2}[y - m_y]^T R_y^{-1}[y - m_y]\} \\
&= [(2\pi)^{(n)/2}|R_x|^{1/2}]^{-1} exp\{-\tfrac{1}{2}[x - m_x]^T R_x^{-1}[x - m_x]\} \\
&\quad .exp\{-\tfrac{1}{2}[y - m_y]^T R_y^{-1}[y - m_y]\} \\
&= p_X(x)p_Y(y)
\end{aligned}
$$

The theorem is proven as $p_Z(z) = p_{X,Y}(x, y)$.

■

Conditional pdf for jointly Gaussian vectors/Minimum-variance Estimation for Jointly Gaussian Variable

Theorem 19. *Suppose X and Y are jointly Gaussian random vectors that are n and m dimensional respectively. with the joint distribution given as*

$$
\begin{aligned}
p_{X,Y}(x,y) &= [(2\pi)^{(n+m)/2}|R_z|^{1/2}]^{-1} \\
&\quad .exp\{-\frac{1}{2}\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}^T R_z^{-1}\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}\}.
\end{aligned}
$$

where

$$R_z = \begin{bmatrix} R_x & R_{xy} \\ R_{yx} & R_y \end{bmatrix}$$

Then the conditional pdf

$$p_{X|Y}(x|y) = \frac{1}{\sqrt{(2\pi)^n |R_{x|y}|}} exp\{-\frac{1}{2}[x - m_{x|y}]^T R_{x|y}^{-1} [x - m_{x|y}]\}$$

where

$$m_{x|y} = m_x + R_{xy} R_{yy}^{-1} (y - m_y) \text{ and } R_{x|y} = R_x - R_{xy} R_y^{-1} R_{yx}.$$

Also the solution to the problem

$$m_x + R_{xy} R_{yy}^{-1} (y - m_y) = arg\{\inf_f [E(X - f(Y))^T (X - f(Y)]\} \qquad (2)$$

Proof: Note that

$$\begin{aligned} R_z^{-1} &= \begin{pmatrix} R_x & R_{xy} \\ R_{xy}^T & R_y \end{pmatrix}^{-1} \\ &= \underbrace{\begin{pmatrix} I & 0 \\ -R_y^{-1}R_{yx} & I \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} \Delta^{-1} & 0 \\ 0 & R_y^{-1} \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} I & -R_{xy}R_y^{-1} \\ 0 & I \end{pmatrix}}_{T}. \end{aligned}$$

where $\Delta = R_x - R_{xy}R_y^{-1}R_{yx}$ is the Schur complement of R_y in R.

It also follows that

$$
\begin{aligned}
|R_z|^{-1} &= \\
&= |R_z^{-1}| \\
&= |L| \left| \begin{pmatrix} \Delta^{-1} & 0 \\ 0 & R_y^{-1} \end{pmatrix} \right| |T| \\
&= | \begin{pmatrix} \Delta^{-1} & 0 \\ 0 & R_y^{-1} \end{pmatrix} | \\
&= |\Delta^{-1}| \; |R_y^{-1}| \\
&= |\Delta|^{-1} \; |R_y|^{-1}
\end{aligned}
$$

From Thoerem 16 it follows that

$$p_Y(y) = \frac{1}{\sqrt{(2\pi)^m |R_y|}} exp\{-\frac{1}{2}[y - m_y]^T R_y^{-1}[y - m_y]\}$$

and from Bayes rule we have

$$
\begin{aligned}
& p_{X|Y}(x|y) \\
= \; & \frac{p_{x,y}(x,y)}{p_Y(y)} \\
= \; & \frac{\sqrt{(2\pi)^m |R_y|}}{(\sqrt{2\pi})^{(n+m)} |R_z|} \\
& .exp\{-\frac{1}{2}\left[\begin{array}{c} x-m_x \\ y-m_y \end{array}\right]^T R_z^{-1}\left[\begin{array}{c} x-m_x \\ y-m_y \end{array}\right] + \frac{1}{2}[y-m_y]^T R_y^{-1}[y-m_y]\} \\
= \; & \frac{\sqrt{(2\pi)^m |R_y|}}{(\sqrt{2\pi})^{(n+m)} |\Delta||R_y|} \\
& .exp\{-\frac{1}{2}\left[\begin{array}{c} x-m_x \\ y-m_y \end{array}\right]^T LDT \left[\begin{array}{c} x-m_x \\ y-m_y \end{array}\right] + \frac{1}{2}[y-m_y]^T R_y^{-1}[y-m_y]\} \\
= \; & \frac{1}{(\sqrt{2\pi})^n |\Delta|} \\
& .exp\{-\frac{1}{2}[x^T - (m_x + R_{xy}R_y^{-1}(y-m_y))^T]\Delta^{-1}[x-(m_x+R_{xy}R_y^{-1}(y-m_y))]\}
\end{aligned}
$$

This proves the first part of the theorem. Equation 2 follows from Theorem 9 and the fact that $E(X|Y) = m_{x|y}$. $\blacksquare$

Estimation with static linear Gaussian models

Theorem 20. *Suppose x and v are n and m dimensional random vectors respectively that are jointly Gaussian with means m_x and 0 respectively and covariances P_x and R_v that are uncorrelated. Suppose*

$$z = Hx + v$$

where $H \in R^{m\times n}$.

Then the conditional pdf $p_{x|z}$ is Gaussian with mean

$$m_{x|z} = m_x + [P_x H^T][HP_xH^T + R_v]^{-1}[z - Hm_x]$$

and covariance

$$P_{x|z} = P_x - [P_x H^T][HP_xH^T + R_v]^{-1}HP_x.$$

Note that

$$m_{x|z} = m_x + K[z - Hm_x] \text{ *and* } P_{x|z} = P_x - KHP_x.$$

Proof: Note that as x and v are jointly Gaussian, it follows that $\begin{bmatrix} x \\ v \end{bmatrix}$ is Gaussian. As any linear transformation of a Gaussian vector is Gaussian it follows that

$$w = \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix}$$

is Gaussian with mean

$$m_w = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} m_x \\ 0 \end{bmatrix} = \begin{bmatrix} m_x \\ Hm_x \end{bmatrix}$$

with covariance

$$R_w = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} P_x & 0 \\ 0 & R_v \end{bmatrix} \begin{bmatrix} I & H^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} P_x & P_x H^T \\ HP_x & HP_x H^T + R_v \end{bmatrix}.$$

From Theorem 19 it follows that the pdf $p_{x|z}(x|z)$ has mean

$$\begin{aligned} m_{x|z} &= m_x + R_{xz}R_z^{-1}(z - m_z) \\ &= m_x + P_x H^T(HP_xH^T + R_v)^{-1}(z - Hm_x) \end{aligned}$$

and the variance

$$\begin{aligned} R_{x|z} &= R_x + R_{xz}R_{zz}^{-1}R_{zx} \\ &= P_x + P_x H^T HP_xH^T + R_v)^{-1}HP_x \end{aligned} .$$

This proves the theorem. ■

Minimum-variance Linear Estimate of X based on Y.

Theorem 21. *Let X and Y be random vectors with dimensions m and n respectively. Let*

$$\mu = \inf_{\hat{X}=KY} E[\|X - \hat{X}\|_2^2], \tag{3}$$

where $\|.\|_2$ is the two norm in R^m. Further assume that $R_{yy} = E(YY^T)$ is invertible. Then the minimizing solution to the optimization problem (3) is given by

$$\hat{X}^0 = E(XY^T)E(YY^T)^{-1}Y = R_{xy}R_y^{-1}Y.$$

Proof: Note that

$$\begin{aligned} \mu \quad = \quad & \inf \textstyle\sum_{i=1}^{m} E[(X_i - \hat{X}_i)^2] \\ \text{subject to} \quad & \\ & \hat{X}_i = k_i^T Y, i = 1, \ldots, m \end{aligned}$$

where

$$K = \begin{bmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_m^T \end{bmatrix}.$$

Note that the above problem can be decoupled into m optimization problems. The optimization variables $\hat{X}_i$ and k_i do not influence the terms in the objective or the constraints on $\hat{X}_j$ and k_j if $i \neq j$. Motivated by this we define

$$\begin{aligned} \mu_i \quad = \quad & \inf \sqrt{E[(X_i - \hat{X}_i)^2]} \\ \text{subject to} \quad & \\ & \hat{X}_i = k_i^T Y. \end{aligned}$$

where the optimization variables are k_i and $\hat{X}_i$. Let

$$M = \text{ span } \{Y_1, Y_2, \ldots Y_n\}.$$

Let

$$Y := [Y_1, \ldots, Y_n]^T.$$

Thus M is a finite dimensional subspace of H the Hilbert space of scalar random variables. Note that

$$\begin{array}{rcl} \mu_i & = & \inf \|X - \hat{X}_i\|_H \\ & \text{subject to} & \\ & & \hat{X}_i \in M. \end{array}$$

where

$$\|Z\|_H = \sqrt{E(Z^2)}.$$

From the classical projection theorem (Theorem 4) it follows that there exists $\hat{X}^0 \in M$ that achieves the minimum and $(X - \hat{X}^0) \in M^\perp$. Therefore there exists scalars α_i^ℓ such that

$$\hat{X}_i^0 = \sum_{\ell=1}^{n} \alpha_i^\ell Y_\ell$$

with

$$\begin{aligned}
< X_i - \hat{X}_i, Y_j >_H &= 0 \text{ for all } j = 1, \ldots, n \\
\Rightarrow \quad < \hat{X}_i, Y_j >_H &= < X_i, Y_j >_H \text{ for all } j = 1, \ldots, n \\
\Rightarrow \quad < \textstyle\sum_{\ell=1}^{m} \alpha_i^{\ell} Y_\ell, Y_j >_H &= < X_i, Y_j >_H \text{ for all } j = 1, \ldots, n \\
\Rightarrow \quad \textstyle\sum_{\ell=1}^{m} (\alpha_i^{\ell} < Y_\ell, Y_j >_H) &= < X_i, Y_j >_H \text{ for all } j = 1, \ldots, n.
\end{aligned}$$

Recasting above in a matrix form we have

$$\underbrace{\begin{pmatrix} < Y_1, Y_1 > & < Y_2, Y_1 > & \cdots & < Y_m, Y_1 > \\ < Y_1, Y_2 > & < Y_2, Y_2 > & \cdots & < Y_m, Y_2 > \\ \vdots & \vdots & \ddots & \vdots \\ < Y_1, Y_m > & < Y_2, Y_m > & \cdots & < Y_m, Y_m > \end{pmatrix}}_{E(YY^T)} \underbrace{\begin{pmatrix} \alpha_i^1 \\ \alpha_i^2 \\ \vdots \\ \alpha_i^m \end{pmatrix}}_{:=\alpha_i} = \underbrace{\begin{pmatrix} < X_i, y_1 >_H \\ < X_i, y_2 >_H \\ \vdots \\ < X_i, y_m >_H \end{pmatrix}}_{=E(X_iY)}.$$

Thus

$$\alpha_i = [E(YY^T)]^{-1} E(X_i Y)$$

and

$$\hat{X}_i^0 = [\alpha_i^1, \ldots, \alpha_i^m] Y = \alpha_i^T Y = E(X_i Y^T)[E(YY^T)]^{-1} Y.$$

This implies that

$$\hat{X}^0 = \begin{pmatrix} E(X_1 Y^T)[E(YY^T)]^{-1}Y \\ \vdots \\ E(X_n Y^T)[E(YY^T)]^{-1}Y \end{pmatrix} = E(XY^T)[E(YY^T)]^{-1}Y.$$

Thus

$$\hat{X}^0 = R_{xy}R_y^{-1}Y.$$

This proves the theorem. ∎

Corollary 1. *$\hat{X}^0$ is the minimum variance linear estimate (mvle) of X based on Y derived in Theorem 21 if and only if*

$$E[(X - \hat{X}^0)Y^T] = 0.$$

Proof: Note that each element $\hat{X}_i^0$ of the $\hat{X}^0$ belongs to M where $M =$ span $\{Y_1, \ldots, Y_m\}$ whereas $(\hat{X}_i^0 - X_i) \in M^\perp$. Therefore it follows from

Theorem 4 that $\hat{X}_i^0$ is a mvle if and only if

$E[(\hat{X}_i^0 - X_i)Y_j] = < \hat{X}_i^0 - X_i, Y_j >= 0$ for all $i = 1, \ldots, m$ and for all $j = 1, \ldots, m$.

Therefore $\hat{X}_i^0$ is a mvle if and only if

$$E[(X - \hat{X}^0)Y^T] = 0.$$

■

Properties of Minimum-variance Linear Estimates

Theorem 22. *The following statements hold:*

1. *The minimum variance linear estimate of* ΓX *with* $\Gamma \in R^{p\times n}$ *based on* Y *is* $\Gamma\hat{X}^0$ *where* $\hat{X}^0$ *is the minimum variance linear estimate of* X *based on* Y.

2. *Let* $\hat{X}^0$ *be the minimum variance linear estimate of* X *based on* Y. *Then* $P^{\frac{1}{2}}\hat{X}^0$ *is the linear estimate based on* Y *minimizing* $E[(\hat{X}-X)^T P(\hat{X}-X)]$ *where* $P \in R^{n\times n}$ *is any positive definite matrix.*

Proof: (1) Let $Z = \Gamma X$. Note that

$$E[(Z-\Gamma\hat{X}^0)Y^T] = E[\Gamma(X-\hat{X}^0)Y^T] = \Gamma E[(X-\hat{X}^0)Y^T] = 0.$$

From Corollary 1 it follows that $\Gamma\hat{X}^0$ is the minimum variance linear estimate of $Z = \Gamma X$ based on Y.

(2) Note that

$$
\begin{aligned}
\mu &= \inf_{\hat{X}=KY} E[(X-\hat{X})^T P(X-\hat{X})] \\
&= \inf_{\hat{X}=KY} E[(P^{\frac{1}{2}}X - P^{\frac{1}{2}}\hat{X})^T (P^{\frac{1}{2}}X - P^{\frac{1}{2}}\hat{X})] \\
&= \inf_{\hat{X}'=P^{\frac{1}{2}}KY} E[(P^{\frac{1}{2}}X - \hat{X}')^T (P^{\frac{1}{2}}X - \hat{X}')] \\
&= \inf_{\hat{X}'=KY} E[(P^{\frac{1}{2}}X - \hat{X}')^T (P^{\frac{1}{2}}X - \hat{X}')]
\end{aligned}
$$

The above problem is to obtain the minimum variance linear estimate of $P^{\frac{1}{2}}X$ which from part (1) of the theorem is given by $P^{\frac{1}{2}}\hat{X}^0$.

This proves the theorem. $\blacksquare$

Minimum-variance Linear Estimate

Theorem 23. *Let*

$$y = W\beta + \epsilon$$

where

- $W \in R^{m \times n}$ *is a known matrix*
- β *is an n-dimensional random vector with* $E(\beta\beta^T) = R \geq 0.$
- *The vector* ϵ *is a* m *dimensional random vector with* $E(\epsilon\epsilon^T) = Q \geq 0.$
- $E(\epsilon\beta^T) = 0.$

Then the minimum variance linear estimate of β *based on* y *is given by*

$$\hat{\beta} = RW^T(WRW^T + Q)^{-1}y = (W^TQ^{-1}W + R^{-1})^{-1}W^TQ^{-1}y$$

and the error covariance is given by

$$P := E[(\beta - \hat{\beta})(\beta - \hat{\beta})^T] = R - RW^T(WRW^T + Q)^{-1}WR = (W^T Q^{-1} W + R^{-1})^{-1}.$$

Note that

$$P^{-1}\hat{\beta} = W^T Q^{-1} y$$

that does not depend on R *the covariance of* β.

Proof: From Theorem 21 we have

$$\hat{\beta} = E(\beta y^T) E(y y^T)^{-1} y$$

with

$$
\begin{aligned}
E(\beta y^T) &= E[\beta(\beta^T W^T + \epsilon^T)] \\
&= E[\beta\beta^T W^T] + E[\beta\epsilon^T] \\
&= RW^T \\
E(yy^T) &= E[(W\beta + \epsilon)(W\beta + \epsilon)^T] \\
&= E[W\beta\beta^T + W\beta\epsilon^T + \epsilon\beta^T W^T + \epsilon\epsilon^T] \\
&= WE[\beta\beta^T]W^T + E[\epsilon\epsilon^T] = WRW^T + Q
\end{aligned}
$$

Therefore we have

$$\hat{\beta} = RW^T(WRW^T + Q)^{-1}y.$$

Note that one can show that

$$RW^T(WRW^T + Q)^{-1} = (W^TQ^{-1}W + R^{-1})^{-1}W^TQ^{-1}$$

by pre-multiplying by $(W^TQ^{-1}W + R^{-1})$ and postmultiplying by $(WRW^T + Q)^{-1}$ that shows that $\hat{\beta} = (W^TQ^{-1}W + R^{-1})^{-1}W^TQ^{-1}y$.

The error covariance matrix is given by

$$
\begin{aligned}
E[(\beta-\hat{\beta})(\beta-\hat{\beta})^T] &= E[\beta(\beta-\hat{\beta})^T] - E[\hat{\beta}(\beta-\hat{\beta})^T] \\
&= E[\beta(\beta-\hat{\beta})^T], \text{ because } \hat{\beta}_i \in M \text{ and } (\beta_j - \hat{\beta}_j) \in M^\perp \\
&= E[\beta\beta^T] - E[\beta\hat{\beta}^T] = E[\beta\beta^T] - E[(\beta-\hat{\beta}+\hat{\beta})\hat{\beta}^T] \\
&= E[\beta\beta^T] - E[(\beta-\hat{\beta})\hat{\beta}^T] - E[\hat{\beta}\hat{\beta}^T] \\
&= E[\beta\beta^T] - 0 - E[\hat{\beta}\hat{\beta}^T] \\
&= R - RW^T(WRW^T+Q)^{-1}WR \\
&= (W^TQ^{-1}W + R^{-1})^{-1}.
\end{aligned}
$$

The last identity follows by using the matrix identity

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{22}A_{11}^{-1}$$

and identifying $A_{11} = R^{-1},\ A_{12} = -W^T,\ A_{21} = W,\ A_{22} = Q$. This proves the theorem. ∎

Combining Estimators

Theorem 24. *Suppose*

$$y_a = W_a\beta + v_a \text{ and } y_b = W_a\beta + v_b$$

where v_a *and* v_b *are uncorrelated (*$E(v_a v_b^T) = 0$*) and* $E[v_a v_a^T] = Q_a,$ *and* $E[v_b v_b^T] = Q_b$*. Suppose* β_a *is the mvle of* β *based on* y_a *and* β_b *is the mvle of* β *based on* y_b *with error covariance matrices* P_a *and* P_b *respectively. Then the mvle of* β *based on*

$$y := \begin{bmatrix} y_a \\ y_b \end{bmatrix}$$

is given by

$$P^{-1}\hat{\beta} = P_a^{-1}\beta_a + P_b^{-1}\beta_b$$

where

$$P^{-1} = P_a^{-1} + P_b^{-1} - R^{-1}$$

with P the associated error covariance matrix of $b\hat{e}ta$..

Proof:

Note that from Theorem 23 it follows that

$$P_a^{-1}\beta_a = W_a^T Q_a^{-1} y_a \text{ and } P_b^{-1}\beta_b = W_b^T Q_b^{-1} y_b$$

with

$$P_a^{-1} = W_a^T Q_a^{-1} W_a + R^{-1} \text{ and } P_b^{-1} = W_b^T Q_b^{-1} W_b + R^{-1}.$$

Its also follows that

$$P^{-1}\hat{\beta} = W^T Q^{-1} y = \begin{bmatrix} W_a^T & W_b^T \end{bmatrix} \begin{bmatrix} Q_a^{-1} & 0 \\ 0 & Q_b^{-1} \end{bmatrix} \begin{bmatrix} y_a \\ y_b \end{bmatrix}.$$

Therefore

$$
\begin{aligned}
P^{-1}\hat{\beta} &= W_a^T Q_a^{-1} y_a + W_b^T Q_b^{-1} y_b \\
&= P_a^{-1}\beta_a + P_b^{-1}\beta_b.
\end{aligned}
$$

Note that it follows from Theorem 23 that

$$
\begin{aligned}
P^{-1} &= W^T Q^{-1} W + R^{-1} \\
&= \begin{bmatrix} W_a^T & W_b^T \end{bmatrix} \begin{bmatrix} Q_a^{-1} & 0 \\ 0 & Q_b^{-1} \end{bmatrix} \begin{bmatrix} W_a \\ W_b \end{bmatrix} \\
&= (W_a^T Q_a^{-1} W_a + R^{-1}) + (W_b^T Q_b^{-1} W_b + R^{-1}) - R^{-1} \\
&= P_a^{-1} + P_b^{-1} - R^{-1}.
\end{aligned}
$$

This proves the theorem.

■

Remark: Note that the new estimate is a linear combination of the estimates β_a and β_b with the weights proportional to the inverse of the error covariance matrices; if P_a is large compared to P_b then β_a contributes lesser towards the new estimate when compared to β_b.

Minimum Variance Linear Unbiased Estimator (Gauss-Markov Estimators)

Let ϵ be a m dimensional random vector with $E(\epsilon\epsilon^T) = Q$ and $E[\epsilon) = 0$. Further let

$$y = W\beta + \epsilon$$

where $\beta \in R^n$ and $W \in R^{m\times n}$. Note that β is a deterministic quantity that is unknown. What is desired is an linear estimator $\hat{\beta}$ of β based on y such that $E[\hat{\beta}] = \beta$. The property of the estimator that $E[\hat{\beta}] = \beta$ is termed as the *unbiased* property.

Note that as $\hat{\beta}$ has to be linear it has the form Ky for some $K \in R^{n\times m}$. Also note the following

$$
\begin{array}{rlcl}
 & E[\hat{\beta}] & = & \beta \text{ for all } \beta \in R^n \\
\Leftrightarrow & E[Ky] & = & \beta \text{ for all } \beta \in R^n \\
\Leftrightarrow & E[KW\beta + \epsilon] & = & \beta \text{ for all } \beta \in R^n \\
\Leftrightarrow & KW\beta + E[\epsilon] & = & \beta \text{ for all } \beta \in R^n \\
\Leftrightarrow & KW\beta & = & \beta \text{ for all } \beta \in R^n \\
\Leftrightarrow & KW & = & I
\end{array}
$$

Note that the performance of the estimator will be measured by the measure

$$
\begin{array}{rcl}
E[\|\beta - \hat{\beta}\|_2^2] & = & E[\|\beta - Ky\|_2^2] \\
 & = & E[\|\beta - KW\beta - K\epsilon\|_2^2] \\
 & = & E[\|K\epsilon\|_2^2 \\
 & = & E[\epsilon^T K^T K\epsilon] \\
 & = & E[Trace[K\epsilon\epsilon^T K^T]] \\
 & = & Trace[KE[\epsilon\epsilon^T]K^T] \\
 & = & Trace[KQK^T]
\end{array}
$$

where we have used the constraint on K that $KW = I$ and the fact that $x^T x = Trace[xx^T]$.

Thus we are interested in the problem outlined in the following theorem

Theorem 25. *Consider the following problem:*

$$\mu = \inf\{Trace[KQK^T] | KW = I\}.$$

The solution to the above problem is given by

$$K_o^T = Q^{-1}W(W^T Q^{-1} W)^{-1}.$$

Proof: Let

$$K^T = [k_1 \; k_2 \; \ldots \; k_n]$$

where k_i is the i^{th} column of K^T. Note that

$$K = \begin{bmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_n^T \end{bmatrix}$$

Note that

$$Trace[KQK^T] = \sum_{i=1}^{n} (KQK^T)_{ii} = \sum_{i=1}^{n} k_i^T Q k_i.$$

Also

$$
\begin{array}{rlcl}
& KW & = & [Kw_1 \ Kw_2 \ \ldots \ KW_n] = I \\
\Leftrightarrow & Kw_i & = & e_i \text{ for all } i = 1 \ldots n \\
\Leftrightarrow & \begin{bmatrix} k_1^T w_i \\ k_2^T w_i \\ \vdots \\ k_n^T w_i \end{bmatrix} & = & e_i \text{ for all } i = 1 \ldots n \\
\Leftrightarrow & k_i^T w_j & = & \delta_{ij} \text{ for all } i = 1, \ldots, n \text{ and } j = 1, \ldots, n
\end{array}
$$

where e_i is the i^{th} unit vector.

Thus the optimization problem can be written as

$$
\mu = \inf\{\sum_{i=1}^{n} k_i^T Q k_i | k_i^T w_j = \delta_{ij} \text{ for all } i, j = 1, \ldots, n\}.
$$

As the optimization over k_i is decoupled from that of k_j the solution to the above problem can be found by solving

$$\mu_i = \inf\{k_i^T Q k_i | k_i^T w_j = \delta_{ij} \text{ for all } j = 1, \ldots, n\}.$$

We define the inner product on R^n as

$$< \alpha, \gamma >= \alpha^T Q \gamma.$$

In the above inner product the solution to μ_i can be found by solving the problem

$$\{\inf\{\|k_i\| \ \mid \ < k_i, Q^{-1} w_j >= \delta_{ij} \text{ for all } j = 1, \ldots, n\}.$$

Appealing to Theorem 8 it follows that there exists an optimal solution $k_{o,i}$ to the above optimization such that $k_{o,i} \in span\{Q^{-1} w_j\}$ with $< k_{o,i}, Q^{-1} w_j >= \delta_{ij}$ for all $j = 1, \ldots, n$.

Thus there exist constants $\alpha_{i,l}$ for all $l = 1, \ldots, n$ such that

$$k_{o,i} = \sum_{l=1}^{n} \alpha_{i,l} Q^{-1} w_l$$

such that

$$< \sum_{l=1}^{n} \alpha_{i,l} Q^{-1} w_l, Q^{-1} w_j >= \delta_{ij} \text{ for all } j = 1, \ldots, n.$$

] that is

$$\sum_{l=1}^{n} [w_j^T Q^{-1} Q Q^{-1} w_l] \alpha_{i,l} = \delta_{ij} \text{ for all } j = 1, \ldots, n.$$

Thus

$$\sum_{l=1}^{n}[w_j^T Q^{-1} w_l]\alpha_{i,l} = \delta_{ij} \text{ for all } j = 1, \ldots, n.$$

Thus

$$w_j^T Q^{-1} W \alpha_i = \delta_{ij} \text{ for all } j = 1, \ldots, n$$

where α_i is a column vector. This implies that

$$W^T Q^{-1} W \alpha_i = e_i$$

and therefore

$$\alpha_i = (W^T Q^{-1} W)^{-1} e_i.$$

Note that

$$k_{o,i} = Q^{-1} W \alpha_i$$

and therefore

$$k_{o,i} = Q^{-1} W (W^T Q^{-1} W)^{-1} e_i \text{ for all } i$$

and therefore

$$K_o^T = [k_{o,1} \; \ldots k_{o,m}] = Q^{-1}W(W^TQ^{-1}W)^{-1}.$$

Thus

The following theorem summarizes the discussion on minimum variance unbiased estimators (mvue).

Theorem 26. *Let*

$$y = W\beta + \epsilon$$

where $\beta \in R^n$, $W \in R^{m\times n}$, $E(\epsilon) = 0$ *and* $E(\epsilon\epsilon^T) = Q$ *where* Q *is invertible. Then the minimum variance unbiased linear estimate of* β *based on* y *is given by* $b\hat{e}ta$ *that satisfies*

$$\hat{\beta} = (W^TQ^{-1}W)^{-1}W^TQ^{-1}y$$

and

$$E[(\beta - \hat{\beta})(\beta - \hat{\beta})^T] = (W^TQ^{-1}W)^{-1}.$$

Proof: Note that the Gauss-markov estimator is given by

$$\hat{\beta} = K_o y = (W^T Q^{-1} W)^{-1} W^T Q^{-1}$$

with K_o defined in the proof of Theorem 25. Also

$$\begin{aligned} E[(\beta - \hat{\beta})(\beta - \hat{\beta})^T] &= E[(\beta - K_o y)(\beta - K_o y)^T] \\ &= E[K_o \epsilon \epsilon^T K_o^T] \\ &= (W^T Q^{-1} W)^{-1} W^T Q^{-1} Q (Q^{-1} W (W^T Q^{-1} W)^{-1}) \\ &= (W^T Q^{-1} W)^{-1} \end{aligned}$$

This proves the theorem.

■

The State Space Model

The following is the state space model of the system.

$$
\begin{aligned}
x_{i+1} &= F_i x_i + G_i u_i \\
y_i &= H_i x_i + v_i
\end{aligned} \tag{4}
$$

We will assume that

- u_i and u_j are not correlated if $i \neq j$.
- v_i and v_j are not correlated if $i \neq j$.
- u_i and v_j are not correlated if $i \neq j$.
- x_0 is not correlated with u_i and v_j.

The statistics are further described by the following equation:

$$E\left(\begin{bmatrix} u_i \\ v_i \\ x_0 \end{bmatrix}, \begin{bmatrix} u_j \\ v_j \\ x_0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} Q_i\delta_{ij} & S_i\delta_{ij} & 0 & 0 \\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix} \tag{5}$$

<u>Comments</u>

- x_j, y_j, u_j and v_j are assumed to be n, m, p and m dimensional random vectors.

- H denotes the Hilbert space of random variables with the inner product defined as $<\alpha, \beta>= E(\alpha\beta)$. This is the only Hilbert space on stochastic entities we will employ.

- By any span of random vectors (possibly of different dimensions) is meant a *component wise* span which is a subspace of the *scalar* Hilbert space H. For example:

$$\text{span}\,\{x_0, u_1, \ldots, u_{j-1}\}$$

is defined to be

$$\text{span}\,\{x_0(1), \ldots, x_0(n), u_1(1), \ldots, u_1(p), u_2(1), \ldots, u_(j-1)(p)\} \subset H.$$

Properties of the Model

Lemma 11. *For the model described by Equations (4) and (5) the following hold:*

1. *(Uncorrelatedness property of states and inputs)*

$$\left.\begin{array}{rcl} E(\, u_i\, {x_j}^* \,) & = & 0 \\ E(\, v_i\, {x_j}^* \,) & = & 0 \end{array}\right\} j \leq i \tag{6}$$

2. *(Uncorrelatedness property of outputs and inputs)*

$$\left.\begin{array}{rcl} E(\, u_i\, {y_j}^* \,) & = & 0 \\ E(\, v_i\, {y_j}^* \,) & = & 0 \end{array}\right\} j \leq i-1 : \tag{7}$$

$$\begin{array}{rcl} E(\, u_i\, {y_i}^* \,) & = & S_i \\ E(\, v_i\, {y_i}^* \,) & = & R_i \end{array} \tag{8}$$

Proof:

(1) Let $j \leq i$. From Equation (4) it follows that

$$x_j \in \text{Span}\{x_0, u_1, \ldots, u_{j-1}\};$$

Note that x_0 is uncorrelated with u_i and u_i is uncorrelated with $u_1, u_2 \ldots u_{j-1}$ (see (5)). (note that : Initial state and inputs are of different dimensions, here span means component wise span; see the comment earlier) and therefore x_j is uncorrelated to u_i that is $E(u_i x_j^*) = 0$.

Also v_i is uncorrelated with x_0 and u_i are correlated only for same time index (see (5)). $E(\ v_i\ x_j{}^*\ ) = 0$ for j ≤i an thus $E(\ v_i\ x_j{}^*\ ) = 0$.

(2) Again note that

$$x_j \in \text{Span}\{x_0, u_1, u_2, \ldots, u_{j-1}\}.$$

$E(\ x_0\ {v_i}^*\ ) = 0$ for all i, $E(\ u_j\ {v_i}^*\ ) = 0$ for all $i \neq j$, and $y_j = H_j x_j + v_j$. Thus if $j \leq i-1$ then $E(u_i y_j^*) = 0$ and $E(v_i, y_j^*) = 0$.

Also note that

$$\begin{aligned}
E(\ u_i\ {y_i}^*\ ) &= E(\ u_i\ (H_i x_i)^*\ ) + E(\ u_i\ {v_i}^*\ ) \\
&= E(\ u_i\ x_i^* {H_i}^*\ ) + S_i \\
&= S_i \\
E(v_i y_i^*) &= E(v_i x_i^*) H_i^* + E(v_i v_i^*) = R_i
\end{aligned}$$

This proves the lemma.

$\blacksquare$

Notation

- By any span of random vectors (possibly of different dimensions) is meant a *component wise* span which is a subspace of the *scalar* Hilbert space H. For example:

  $$\text{span}\ \{x_0, u_1, \ldots, u_{j-1}\}$$

  is defined to be

  $$\text{span}\ \{x_0(1), \ldots, x_0(n), u_1(1), \ldots, u_1(p), u_2(1), \ldots, u_(j-1)(p)\} \subset H.$$

- Suppose M is a closed subspace of the scalar Hilbert space H and $z \in H$ is a random variable. We define the *projection* of z onto M by

  $$Pr_M(z) := arg\{\inf_{m \in M} \|z - m\|\} = arg\{\inf_{m \in M} E[(z-m)^2]\}.$$

  Note that from the classical projection theorem (Theorem 4), $Proj_M(z)$ is guaranteed to exist with the property that

$$\tilde{z} := z - Pr_M(z) \in M^{\perp}.$$

- Suppose z is a q dimensional random vector $z = (z(1), \ldots, z(q))^T$. Then the projection of z onto the closed subspace M is defined as

$$Pr_M(z) := \begin{bmatrix} Pr_M z(1) \\ Pr_M z(2) \\ \vdots \\ Pr_M z(q) \end{bmatrix};$$

that is $P_M(z)$ is the componentwise projection of z onto M.

Associated with this projection is the error vector:

$$\tilde{z} := z - Pr_M(z).$$

It is evident from Theorem 4 that every element of $\tilde{z} \in M^{\perp}$.

- In the context of the model (4). Let $M_j :=$ span $\{y_0, \ldots, y_j\}$. Then the projection of a random vector z onto M_j is denoted by $\hat{z}_{|j}$. That is

$$\hat{z}_{|j} := Pr_{M_j}\, z.$$

Associated with this projection is the error vector:

$$\tilde{z}_{|j} := z - \hat{z}_{|j}.$$

It is evident from Theorem 4 that every element of $\tilde{z}_{|j} \in M_j^{\perp}$.

- Using the above notation we have

  ⋆ $\hat{x}_{i|j} =$ the projection of the state at time instant i onto $M_j =$ span $\{y_0, \ldots, y_j\}$

- ⋆ $\tilde{x}_{i|j}$ = The associated error vector.

- We will also use the notation

$$\hat{x}_i := \hat{x}_{i|i-1}.$$

Thus $\hat{x}_i$ is the projection of x_i onto the past $i-1$ measurements. $\tilde{x}_i$ will denote the associated error vector.

Innovations

Note that

$$\hat{y}_{i|j} = Pr_{M_j}(y_i),$$

$M_j = \{y_0, \ldots, y_j\}$.

The innovation sequence is defined by

$$e_i \triangleq y_i - \hat{y}_{i|i-1}.$$

Theorem 27. *The innovation sequence is white, that is:*

$$E(\ e_i\ {e_j}^*\ ) = 0 \textit{ if } \ i \neq j.$$

Proof: Note that:

$$e_i \in \text{ span } \{y_0, y_1, \ldots, y_i\}.$$

as $e_i = y_i - \hat{y}_{i|i-1}$ and $\hat{y}_{i|i-1} \in M_{i-1} =$ span $\{y_0, \ldots, y_{i-1}\}$.

From classical projection theorem (Theorem 4) it follows that for all $k = 1, 2, \ldots, i-1$

$$E(\ (y_i - \hat{y}_{i|i-1})\ (y_k)^*\ ) = 0.$$

This implies that

$$E(\ e_i\ y_k{}^*\ ) = 0 \text{ for all } k < i$$

which in turn implies that

$$E(\ e_i\ e_k{}^*\ ) = 0 \text{ for all } k < i$$

This proves the theorem. ■

Lemma 12.

$$M_j := \textit{span}\,\{y_0, \ldots, y_j\} = \textit{span}\,\{e_0, \ldots, e_j\}.$$

Proof: Left to the reader. ■

Measurement Model with Innovations as Input

In the discussion below we will obtain a state space representation where the input is the sequence e_i and the output is the measurement sequence y_i. Thus we will develop a model to obtain causally the measurement sequence from the innovation sequence.

Lemma 13. *For the state space model (4) with the statistics described by (5) we have*

$$\begin{aligned} \hat{x}_{i+1|i} &= F_i \hat{x}_{i|i-1} + K_{p_i} e_i \\ y_i &= H_i \hat{x}_{i|i-1} + e_i \end{aligned} \tag{9}$$

where

$$K_{p,i} := E(\, x_{i+1}\, {e_i}^* \,) {R_{e,i}}^{-1}.$$

Proof: From

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i \\ y_i &= H_i x_i + v_i \end{aligned}$$

we have

$$\begin{aligned}
\hat{x}_{i+1|i} &= Pr_{\mathsf{span}(y_0,y_1,\ldots,y_j)}(x_{i+1}) \\
&= Pr_{\mathsf{span}(e_0,e_1,\ldots,e_j)}(x_{i+1}) \\
&= Pr_{\mathsf{span}(e_0)}(x_{i+1}) + \ldots + Pr_{\mathsf{span}(e_i)}(x_{i+1})
\end{aligned}$$

From Theorem 21 we have that

$$Pr_{\mathsf{span}(e_j)}(x_{i+1}) = E(x_{i+1}e_j^*)E(e_je_j^*)^{-1}e_j.$$

Thus

$$
\begin{aligned}
\hat{x}_{i+1|i} &= E(\, x_{i+1}\, {e_0}^* \,)E(\, e_0\, {e_0}^* \,)^{-1} e_0 + \ldots \\
&= \sum_{j=0}^{i} E(\, x_{i+1}\, {e_j}^* \,)E(\, e_j\, {e_j}^* \,)^{-1} e_j \\
&= \sum_{j=0}^{i-1} E(\, x_{i+1}\, {e_j}^* \,)E(\, e_j\, {e_j}^* \,)^{-1} e_j + E(\, x_{i+1}\, {e_i}^* \,){R_{e,i}}^{-1} e_i \\
&= \hat{x}_{i+1|i-1} + \underbrace{E(\, x_{i+1}\, {e_i}^* \,){R_{e,i}}^{-1}}_{K_{p,i}} e_i
\end{aligned}
$$

Also it follows from Theorem 22 and Lemma 11 that

$$\begin{aligned}
\hat{x}_{i+1}|_{i-1} &= (\widehat{F_i x_i})_{|i-1} + (\widehat{G_i u_i})_{|i-1} \\
&= F_i \hat{x}_{i|i-1} + G_i u_{1|i-1} \\
&= F_i \hat{x}_{i|i-1}
\end{aligned}$$

Thus it follows that

$$\begin{aligned}
\hat{x}_{i+1|i} &= F_i \hat{x}_{i|i-1} + K_{p_i} e_i \\
y_i &= \hat{y}_{i|i-1} + e_i = H_i \hat{x}_{i|i-1} + e_i
\end{aligned}$$

This proves the lemma. ■

Innovation Model with Measurements as Input

Now we obtain a state space representation where the input is the sequence y_i and the output is the innovation sequence e_i. Thus we will obtain a state space model to obtain causally the innovation sequence from the measurements.

Lemma 14. *For the state space model (4) with the statistics described by (5) we have*

$$\begin{aligned} \hat{x}_{i+1|i} &= F_{p,i}\hat{x}_{i|i-1} + K_{p_i}y_i \\ e_i &= -H_i\hat{x}_{i|i-1} + y_i \end{aligned} \qquad (10)$$

where

$$K_{p,i} := E(\ x_{i+1}\ {e_i}^*\ ){R_{e,i}}^{-1}$$

and

$$F_{p_i} := F_i - K_{p_i}H_i.$$

Proof:

$$\begin{aligned}
\hat{x}_{i+1|i} &= F_i \hat{x}_{i|i-1} + K_{p_i} e_i \\
&= F_i \hat{x}_{i|i-1} + K_{p_i} [y_i - \hat{y}_{i|i-1}] \\
&= F_i \hat{x}_{i|i-1} + K_{p_i} [y_i - H_i \hat{x}_{i|i-1}] \\
&= (F_i - K_{p_i} H_i) \hat{x}_{i|i-1} + K_{p_i} y_i
\end{aligned}$$

Letting

$$F_{p_i} := F_i - K_{p_i} H_i$$

we have

$$\begin{aligned}
\hat{x}_{i+1|i} &= F_{p_i} \hat{x}_{i|i-1} + K_{p_i} y_i \\
e_i &= y_i - \hat{y}_{i|i-1} = y_i - H_i \hat{x}_{i|i-1} \\
&= -H_i \hat{x}_{i|i-1} + y_i
\end{aligned}$$

Note that we have defined $\hat{x}_i := \hat{x}_{i|i-1}$. Thus

$$\begin{aligned} \hat{x}_{i+1} &= F_i \hat{x}_i + K_{p_i} e_i \\ y_i &= H_i \hat{x}_i + e_i \end{aligned}$$

$$\begin{aligned} \hat{x}_{i+1} &= F_{p_i} \hat{x}_i + K_{p_i} y_i \\ e_i &= -H_i \hat{x}_i + y_i \end{aligned}$$

This proves the lemma. ■

Thus if we determine K_{p_i} and R_{e_i} we can obtain a causal and causally invertible model for the process y_i as follows,

$$
\begin{aligned}
\hat{x}_{i+1} &= F_i \hat{x}_i + K_{p_i} e_i \\
y_i &= H_i \hat{x}_i + e_i
\end{aligned}
$$

$$
\begin{aligned}
\hat{x}_{i+1} &= F_{p_i} \hat{x}_i + K_{p_i} y_i \\
e_i &= -H_i \hat{x}_i + y_i
\end{aligned}
$$

Original Model:

$$
\begin{aligned}
x_{i+1} &= F_i x_i + G_i u_i \\
y_i &= H_i x_i + v_i
\end{aligned}
$$

State Recursions

Define

$$\Pi_i \quad := \quad E(\ x_i\ x_i{}^*\ )$$

Using

$$x_{i+1} = F_i x_i + G_i u_i$$

we have

$$\begin{aligned}
\Pi_{i+1} \quad &= \quad E(\ x_{i+1}\ x_{i+1}{}^*\ ) \\
&= \quad F_i E(\ x_{i+1}\ x_{i+1}{}^*\ ) F_i^* + G_i E(\ u_i\ u_i{}^*\ ) G_i^* \\
&\quad\quad F_i E(\ x_i\ u_i{}^*\ ) G_i^* + G_i E(\ u_i\ x_i{}^*\ ) F_i^* \\
&= \quad F_i \Pi_i F_i^* + G_i Q_i G_i^*
\end{aligned}$$

State Estimate Recursions

Define:

$$\Sigma_i := E(\ \hat{x}_i\ \hat{x}_i^*\ )$$

Using Lemma 13 we have

$$\begin{aligned} \Sigma_{i+1} &= F_i \Sigma_i F_i^* + K_{p_i} R_{e_i} K_{p_i}^* \\ \Sigma_0 &= 0 \end{aligned}$$

Error Covariance Recursions

Define

$$\begin{aligned} P_i &= E(\ \tilde{x}_{i|i-1}\ \tilde{x}^*_{i|i-1}\ ) \text{ with} \\ \tilde{x}_{i|i-1} &:= x_i - \hat{x}_{i|i-1} \\ &=: \tilde{x}_i \end{aligned}$$

It follows that

$$\begin{aligned} P_i &= E(\ (x - \hat{x}_i)\ (x - \hat{x}_i)^*\ ) \\ &= E(\ x_i\ {x_i}^*\ ) - E(\ \hat{x}_i\ \hat{x}_i^*\ ) \\ &= \Pi_i - \Sigma_i \\ P_{i+1} &= \Pi_{i+1} - \Sigma_{i+1} \\ &= F_i P_i F_i^* + G_i Q_i G_i^* - K_{p_i} R_{e_i} k^*_{p_i} \end{aligned}$$

K_{p_i} and R_{e_i} in terms of P_i

R_{ei} in terms of P_i: Note that

$$\begin{aligned} R_{e_i} &= E(\ e_i\ e_i{}^*\ ) \\ &= H_i E(\ \tilde{x}_i\ \tilde{x}_i^*\ ) H_i^* + E(\ v_i\ v_i{}^*\ ) \\ &= H_i P_i H_i^* + R_i \end{aligned}$$

K_{pi} in terms of P_i: Note that

$$\begin{aligned} e_i &= y_i - y_{i|i-1} \\ &= H_i x_i + v_i - H_i x_{i|i-1} \\ &= H_i (x_i - x_{i|i-1}) + v_i \\ &= H_i \tilde{x}_i + v_i \end{aligned}$$

Thus

$$
\begin{aligned}
E(\ x_{i+1}\ e_i{}^*\ ) &= F_i E(\ x_i\ e_i{}^*\ ) + G_i E(\ u_i\ e_i{}^*\ ) \\
&= F_i E(\ x_i\ e_i{}^*\ ) + G_i E(\ u_i\ e_i{}^*\ ) \\
&= F_i E(\ x_i\ \tilde{x}_i^*\ ) H_i^* + F_i E(\ x_i\ v_i{}^*\ ) \\
&\quad + G_i E(\ u_i\ \tilde{x}_i^*\ ) H_i^* + G_i E(\ u_i\ v_i{}^*\ ) \\
&= F_i E(\ (\tilde{x}_i + \hat{x}_i)\ \tilde{x}_i^*\ ) H_i^* + G_i E(\ u_i\ \tilde{x}_i^*\ ) H_i^* + G_i S_i \\
&= F_i P_i H_i^* + G_i S_i \text{ because } \tilde{x}_i \perp u_i
\end{aligned}
$$

Thus

$$K_{pi} = E(x_{i+1} e_i^*) R_{e,i}^{-1}$$

$$= (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$$

Innovation Recursions

Theorem 28.

$$
\begin{aligned}
x_{i+1} &= F_i x_i + G_i u_i \\
y_i &= H_i x_i + v_i
\end{aligned}
$$

with the statistics

$$
E\left(\begin{bmatrix} u_i \\ v_i \\ x_0 \end{bmatrix}, \begin{bmatrix} u_j \\ v_j \\ x_0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} Q_i\delta_{ij} & S_i\delta_{ij} & 0 & 0 \\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}
$$

The innovation sequence e_i can be generated recursively as described below:

$$\begin{aligned} e_i &= y_i - \hat{y}_{i|i-1} \\ \hat{x}_0 &= 0 \\ e_0 &= y_0 \end{aligned}$$

$$\begin{aligned} \hat{x}_{i+1|i} &= F_i \hat{x}_{i|i-1} + K_{p_i} e_i \\ K_{p_i} &= \left(F_i P_i F_i^* + G_i S_i \right) R_{e_i} \\ R_{e_i} &= H_i P_i H_i^* + R_i \\ P_{i+1} &= F_i P_i F_i^* + G_i Q_i G_i^* - R_{e_i} K_{p_i}^* \end{aligned}$$

where

$$
\begin{aligned}
P_i &:= E(\, \tilde{x}_i \, \tilde{x}_i^* \,) \\
\tilde{x}_i &= x_i - \hat{x}_i \\
P_0 &= \Pi_0
\end{aligned}
$$

Proof: Follows from the development before. ∎

Filter State Recursions

Theorem 29. *Consider*

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i \\ y_i &= H_i x_i + v_i \end{aligned}$$

with the statistics

$$E\left(\begin{bmatrix} u_i \\ v_i \\ x_0 \end{bmatrix}, \begin{bmatrix} u_j \\ v_j \\ x_0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} Q_i\delta_{ij} & S_i\delta_{ij} & 0 & 0 \\ S_i^*\delta_{ij} & R_i\delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}$$

The filtered sequence $\hat{x}_i$ can be generated recursively as described below:

$$\begin{aligned} \hat{x}_{i+1} &= F_{p,i}\hat{x}_i + K_{p_i} y_i \\ \hat{x}_0 &= 0 \end{aligned}$$

$$
\begin{aligned}
K_{p_i} &= \left(F_i P_i F_i^* + G_i S_i\right) R_{e_i} \\
R_{e_i} &= H_i P_i H_i^* + R_i \\
P_{i+1} &= F_i P_i F_i^* + G_i Q_i G_i^* - R_{e_i} K_{p_i}^*
\end{aligned}
$$

where

$$
\begin{aligned}
P_i &:= E(\ \tilde{x}_i\ \tilde{x}_i^*\ ) \\
\tilde{x}_i &= x_i - \hat{x}_i \\
P_0 &= \Pi_0
\end{aligned}
$$

Proof: Follows from the development before. ■

Time and measurement update formalism for Kalman Filter

Measurement Update Step

Theorem 30. *Consider the state-space model*

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i \\ y_i &= H_i x_i + v_i \end{aligned}$$

where the estimate $\hat{x}_i = \hat{x}_{i|i-1}$ *is available and also the related error-covariance* $P_i := P_{i|i-1} = E[(x_i - x_{i|i-1})(x_i - x_{i|i-1})^*]$*. Suppose a new measurement* y_i *is obtained. Then the updated estimate of* x_i *and the updated error covariance is given by*

$$\begin{aligned} \hat{x}_{i|i} &= \hat{x}_i + K_{f,i} e_i, \quad K_{f,i} := P_i H_i^* R_{e,i}^{-1} \\ P_{i|i} &= \|x_i - \hat{x}_{i|i}\|^2 = P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i. \end{aligned}$$

Proof: Note that

$$\begin{aligned}\hat{x}_{i|i} = Pr_{sp(y_0,\ldots,y_i)} x_i &= Pr_{sp(e_0,\ldots,e_i)} x_i \\ &= Pr_{sp(e_0,\ldots,e_{i-1})} x_i + Pr_{sp(e_i)} x_i \\ &= \hat{x}_{i|i-1} + < x_i, e_i > R_{e,i}^{-1} e_i.\end{aligned}$$

Note that

$$e_i = y_i - \hat{y}_{i|i-1} = Hx_i + v_i - (H\hat{x}_{i|i-1} + v_{i|i-1} = H(x_i - \hat{x}_{i|i-1}) + v_i = H\tilde{x}_i + v_i.$$

$$\begin{aligned}E[x_i e_i^*] &= E[(\tilde{x}_i + \hat{x}_{i|i-1})e_i^*] \\ &= E[\tilde{x}_i e_i^*] + E[\hat{x}_{i|i-1} e_i^*] \\ &= E[\tilde{x}_i (H\tilde{x}_i + v_i)^*] + E[\hat{x}_{i|i-1}(H\tilde{x}_i + v_i)^*] \\ &= E[\tilde{x}_i \tilde{x}_i^*]H^* + E[\tilde{x}_i v_i^*] + E[\hat{x}_{i|i-1}\tilde{x}_i^*]H^* + E[\hat{x}_{i|i-1} v_i^*] \\ &= P_i H^*\end{aligned}$$

where as $\hat{x}_i \in span\{y_0, \ldots, y_{i-1}\}$ and $\tilde{x}_i \in [sp\{y_0, \ldots, y_{i-1}\}]^\perp$ we have $E[\hat{x}_i \tilde{x}_i^*] = 0$ and $E[\tilde{x}_i v_i^*] = E[(x_i - \hat{x}_i) v_i^*] = E[x_i v_i^*] - E[\hat{x}_i v_i^*] = 0$. This proves that

$$\hat{x}_{i|i} \;\; = \;\; \hat{x}_i + K_{f,i} e_i, \;\; K_{f,i} := P_i H_i^* R_{e,i}^{-1}.$$

Note that

$$\begin{aligned} E[(x_i - x_{i|i})(x_i - x_{i|i})^*] &= E[(x_i - \hat{x}_i - K_{f,i} e_i)(x_i - \hat{x}_i - K_{f,i} e_i)^*] \\ &= E[(\tilde{x}_i - K_{f,i} e_i)(\tilde{x}_i - K_{f,i} e_i)^*] \\ &= E[\tilde{x}_i \tilde{x}_i^*] - E[\tilde{x}_i e_i^*] K_{f,i}^* - K_{f,i} E[e_i \tilde{x}_i^*] \\ &\quad + K_{f,i} E[e_i e_i^*] K_{f,i}^* \\ &= P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i \\ &\quad + P_i H_i^* R_{e_i}^{-1} R_{e,i} R_{e,i}^{-1} H_i P_i \\ &= P_i - P_i H_i^* R_{e,i}^{-1} H_i P_i \end{aligned}$$

where we have used

$$E[\tilde{x}_i e_i^*] = E[x_i e_i^*] - E[\hat{x}_i e_i^*] = P_i H_i^* - E[\hat{x}_i (H \tilde{x}_i + v_i)^*] = P_i H_i^*.$$

Time Update Step

Theorem 31. *Consider the state-space model*

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i \\ y_i &= H_i x_i + v_i \end{aligned}$$

where the estimate $\hat{x}_i = \hat{x}_{i|i}$ *and the related error-covariance* $P_{i|i} := E[(x_i - \hat{x}_{i|i})(x_i - \hat{x}_{i|i})^*]$ *are available. Then the estimate of* x_{i+1} *and the error covariance* $P_{i+1|i}$ *can be obtained as*

$$\begin{aligned} \hat{x}_{i+1|i} &= F_i \hat{x}_{i|i} + G_i \hat{u}_{i|i}, \quad \hat{u}_{i|i} := S_i R_{e,i}^{-1} e_i \\ P_{i+1} = P_{i+1|i} &= \|x_{i+1} - \hat{x}_{i+1|i}\|^2 = F_i P_{i|i} F_i^* + G_i (Q_i - S_i R_{e,i}^{-1} S_i^*) G_i^* \\ &\quad - F_i K_{f,i} S_i^* G_i^* - G_i S_i K_{f,i}^* F_i^* \end{aligned}$$

Proof: Note that

$$
\begin{aligned}
\hat{u}_{i|i} &= Pr_{sp(y_0,\ldots,y_i)} u_i \\
&= Pr_{sp(e_0,\ldots,e_i)} u_i \\
&= Pr_{sp(e_0,\ldots,e_{i-i}) u_i + Pr_{e_i}} u_i \\
&= Pr_{sp(y_0,\ldots,y_{i-i})} u_i + E[u_i e_i^*] R_{e,i}^{-1} e_i
\end{aligned}
$$

Note that

$$
\begin{aligned}
E[u_i e_i^*] &= E[u_i (H\tilde{x}_i + v_i)^*] \\
&= [u_i (x_i - \hat{x}_i)^*] = E[u_i v_i^*] \\
&= E[u_i v_i^*] = S_i
\end{aligned}
$$

Note that $E[u_i \hat{x}_i^*] = E[u_i \hat{x}_i^*] = 0$ as $\hat{x}_i \in sp\{y_0, \ldots, y_{i-1}\}$ and x_i and $y_j, \ j < i$ are no correlated with u_i.

The error covariance update can be obtained easily and left to the reader. ∎

- **Proof:** Note that

$$
\begin{aligned}
\hat{\mathbf{x}}_{\mathbf{i|N}} = Pr_{sp(y_0,\ldots,y_N)}x_i &= Pr_{sp(e_0,\ldots,e_N)}x_i \\
&= Pr_{sp(e_0,\ldots,e_{i-1})}x_i + Pr_{sp(e_i,e_{i+1}\cdots e_N)}x_i \\
&= \hat{\mathbf{x}}_{\mathbf{i|i-1}} + \textstyle\sum_{\mathbf{j=i}}^{\mathbf{N}} \mathbf{E}(\mathbf{x_i e_j^*})\mathbf{R_{e,j}^{-1}e_j}.
\end{aligned}
$$

and

$$
e_j = y_j - \hat{y}_{j|j-1} = H_j x_i + v_j - (H_j\hat{x}_{j|j-1} + v_{j|j-1}) = H_j(x_j - \hat{x}_{j|j-1}) + v_j = H_j\tilde{x}_j + v_j.
$$

$$
\begin{aligned}
\mathbf{E}[\mathbf{x_i e_j^*}] &= E[(\tilde{x}_i + \hat{x}_{i|i-1})e_j^*] \\
&= E[\tilde{x}_i e_j^*] + E[\hat{x}_{i|i-1}e_j^*] \\
&= E[\tilde{x}_i(H_j\tilde{x}_j + v_j)^*] + E[\hat{x}_{i|i-1}(H_j\tilde{x}_j + v_j)^*] \\
&= E[\tilde{x}_i\tilde{x}_j^*]H_j^* + E[\tilde{x}_i v_j^*] + E[\hat{x}_{i|i-1}\tilde{x}_j^*]H_j^* + E[\hat{x}_{i|i-1}v_j^*] \\
&= \mathbf{P_{ij}H_j^*}
\end{aligned}
$$

where since $\hat{x}_i \in span\{e_0,\ldots,e_{i-1}\}$ and $\tilde{x}_j \in [sp\{e_0,\ldots,e_{i-1}\}]^\perp$ for $j \geq i$ we have $E[\hat{x}_i\tilde{x}_j^*] = 0$ and $E[\tilde{x}_i v_j^*] = E[(x_i - \hat{x}_i)v_j^*] = E[x_i v_j^*] - E[\hat{x}_i v_j^*] = 0$.

- Also from $x_{i+1} = F_i x_i + G_i u_i$ and $\hat{x}_{i+1} = F_i \hat{x}_i + K_{pi} e_i$ we have

$$\tilde{x}_{i+1} = F_i \tilde{x}_i + G_i u_i - K_{pi} v_i.$$

From this recursion equation, we can derive that for $j \geq i$

$$\begin{aligned} \tilde{x}_j &= \phi_p(j,i)\tilde{x}_i + \sum_{l=i}^{j-1} \phi_p(j-l-1)(G_l u_l - K_{pl} v_l) \\ \Rightarrow \mathbf{P_{ij}} = E(\tilde{x}_i \tilde{x}_j^*) &= E(\tilde{x}_i \tilde{x}_i^*)\phi_p^*(j,i) = \mathbf{P_i \phi_p^*(j,i)} \end{aligned}$$

since $E(\tilde{x}_i \tilde{u}_l^*) = 0$ and $E(\tilde{x}_i \tilde{v}_l^*) = 0$ for $l \geq i$ because $\tilde{x}_i = x_i - \hat{x}_i$ and $\hat{x}_i$ is in $sp(y_0, \cdots, y_{i-1})$. Therefore

$$\hat{x}_{i|N} = \hat{x}_i + P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} e_j$$

Now $\hat{x}_{i|N} = \hat{x}_i + P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} e_j$, Therefore

$$
\begin{aligned}
\tilde{x}_{i|N} &= \tilde{x}_i - P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} e_j \\
\Rightarrow P_{i|N} &= E(\tilde{x}_i \tilde{x}_i^*) - P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} E(e_j \tilde{x}_i^*) - \sum_{j=i}^{N} E(x_i e_j^*) R_{e_j}^{-1} H_j \phi_p(j,i) \\
&\quad + P_i \sum_{j,l=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} E(e_j e_l^*) R_{e_l}^{-1} H_l \phi_p(l,i) P_i \\
&= P_i - 2P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} H_j \phi_p(j,i) P_i + P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} H_j \phi_p(\\
\Rightarrow P_{i|N} &= P_i - P_i \sum_{j=i}^{N} \phi_p^*(j,i) H_j^* R_{e_j}^{-1} H_j \phi_p(j,i) P_i
\end{aligned}
$$

Convex Analysis

- One of the most important concepts in optimization is that of convexity. It can be said that the only true global optimization results involve convexity in one way or another.

- Establishing that a problem is equivalent to a finite dimensional convex optimization is often considered as solving the problem. This viewpoint is further reinforced due to efficient software packages available for convex programming.

Convex Sets

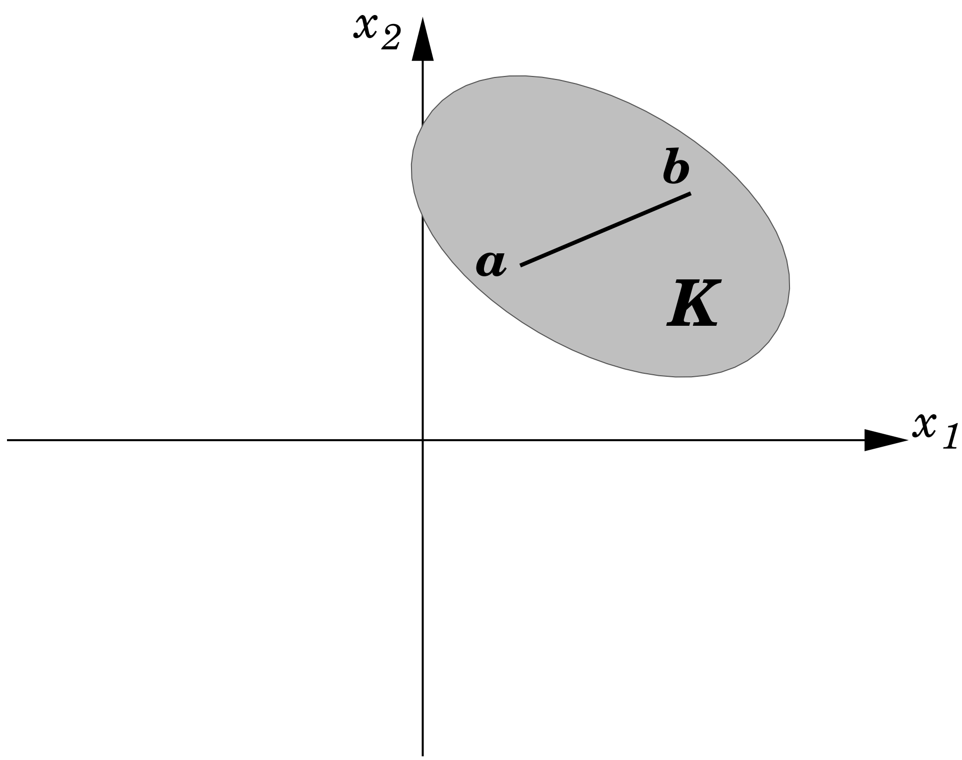

Figure 1: In a convex set a chord joining any two elements of the set lies inside the set.

Definition 23. [Convex sets] *A subset Ω of a vector space X is said to be* *convex if for any two elements* c_1 *and* c_2 *in* Ω *and for a real number* λ *with* $0 < \lambda < 1$ *the element* $\lambda c_1 + (1 - \lambda) c_2 \in \Omega$ *(see Figure 1). The set* $\{\}$ *is assumed to be convex.*

Convex Sets, convex combinations and cones

Definition 24. [Convex combination] *A vector of the form* $\sum_{k=1}^{n} \lambda_k x_k$, *where* $\sum_{k=1}^{n} \lambda_k = 1$ *and* $\lambda_k \geq 0$ *for all* $k = 1, \ldots, n$ *is a convex combination of the vectors* $x_1, \ldots, x_n$.

Definition 25. [Cones] *A subset* C *of a vector space* X *is a cone if for every non-negative* α *in* R *and* c *in* C, $\alpha c \in C$.

A subset C *of a vector space is a convex cone if* C *is convex and is also a cone.*

Definition 26. [Positive cones] *A convex cone* P *in a vector space* X *is a positive convex cone if a relation* $' \geq '$ *is defined on* X *based on* P *such that for elements* x *and* y *in* X, $x \geq y$ *if* $x - y \ \in P$. *We write* $x > 0$ *if* $x \in int(P)$. *Similarly* $x \leq y$ *if* $x - y \ \in -P := N$ *and* $x < 0$ *if* $x \in int(N)$. *Given a vector space* X *with positive cone* P *the positive cone in* X^* *,* $P^{\oplus}$ *is defined as*

$$P^{\oplus} := \{x^* \in X^* :< x, x^* > \geq 0 \text{ for all } x \in P\}.$$

Example 1. *Consider the real number system R. The set*

$$P := \{x : x \text{ is nonnegative}\},$$

defines a cone in R. It also induces a relation $\geq$ on R where for any two elements x and y in R, $x \geq y$ if and only if $x - y \in P$. The convex cone P with the relation $\geq$ defines a positive cone on R.

Minimum Distance to a closed convex set

Theorem 32. *Suppose K is a closed convex subset of a Hilbert space H. Let x be an element in H. Then k_0 satisfies*

$$\|x - k_0\| \leq \|x - k\| \text{ for all } k \in K$$

if and only if

$$< x - k_0, k - k_0 > \leq 0 \text{ for all } k \in K.$$

Proof: ($\Rightarrow$) Suppose $k_0 \in K$ is such that

$$\|x - k_0\| \leq \|x - k\| \text{ for all } k \in K$$

. Then

$$\begin{aligned} < x-k, x-k > &= < x-k_0+k_0-k, x-k_0+k_0-k > \\ &= < x-k_0, x-k_0 > +2 < x-k_0, k_0-k > + < k_0-k, k_0-k \\ &= \|x-k_0\|^2 + \|k-k_0\|^2 - 2 < x-k_0, k-k_0 > \\ &\geq 0. \end{aligned}$$

($\Leftarrow$)Suppose there exists a $k \in K$ such that

$$< x-k_0, k-k_0 >= \epsilon > 0.$$

Let

$$k_\alpha = \alpha k + (1-\alpha)k_0 \text{ and } f(\alpha) := \|x-k_\alpha\|^2.$$

Note that

$$\begin{aligned} f(\alpha) &= < x-\alpha k-(1-\alpha)k_0, x-\alpha k-(1-\alpha)k_0 > \\ &= < x-k_0-\alpha(k-k_0), x-k_0-\alpha(k-k_0) > \\ &= < x-k_0, x-k_0 > -2\alpha < x-k_0, k-k_0 > +\alpha^2 < k-k_0, k-k_0 > \\ &= \|x-k_0\|^2 - 2\alpha\epsilon + \alpha^2\|k-k_0\|^2 \end{aligned}$$

and therefore

$$\left.\frac{df(\alpha)}{d\alpha}\right|_{\alpha=0} = -2\epsilon < 0.$$

Thus in a small neighbourhood of zero $f(\alpha) < f(0)$. Thus there exists a $0 < \alpha < 1$ and a $k_\alpha \in K$ such that

$$\|x-k_\alpha\|^2 = f(\alpha) \leq f(0) = \|x-k_0\|^2.$$

This proves the theorem. ∎

Theorem 33. *Suppose K is a closed convex subset of a Hilbert space H. Let x be an element in H and consider the following optimization problem:*

$$\mu = \inf\{\|x - k\| : k \in K\}.$$

Then there exists a $k_0 \in K$ such that $\|x - k_0\| = \mu$ that is there exists a minimizing solution. Furthermore, k_0 satisfies

$$\|x - k_0\| \leq \|x - k\| \textit{ for all } k \in K$$

if and only if

$$< x - k_0, k - k_0 > \leq 0 \textit{ for all } k \in K.$$

Proof: If $x \in K$, then $k_0 = x$ and theorem is proven. Suppose $x \notin K$. Then for any $n \in N$. there exists $k_n \in M$ such that $\mu \leq \|x - k_n\| \leq \mu + \frac{1}{n}$. Thus there exist a sequence $\{k_i\}_{i=0}^{\infty} \in K$ such that $\|x - k_n\|$ converges to μ as $n \to \infty$. From the parallelogram law, for any integer i and j,

$$\|(k_j - x) + (x - k_i)\|^2 + \|(k_j - x) - (x - k_i)\|^2 = \ 2\|k_j - x\|^2 + 2\|k_i - x\|^2.$$

This implies that

$$\|k_j - k_i\|^2 + \|k_j + k_i - 2x\|^2 = \ 2\|k_j - x\|^2 + 2\|k_i - x\|^2.$$

Thus

$$\|k_j - k_i\|^2 = 2\|k_j - x\|^2 + 2\|k_i - x\|^2 - 4\|\frac{k_j + k_i}{2} - x\|^2. \tag{11}$$

Note that $\frac{k_j + k_i}{2} \in K$, and therefore

$$\|\frac{k_j + k_i}{2} - x\| \geq \ \mu.$$

From (11) we have

$$\|k_j - k_i\|^2 \leq 2\|k_j - x\|^2 + 2\|k_i - x\|^2 - 4\mu^2.$$

Given any $\epsilon > 0$, let N be a large positive integer such that for all $n \geq N$, $\|x - k_n\|^2 \leq \mu^2 + \frac{\epsilon^2}{4}$. If $i, j > N$ then,

$$\|k_j - k_i\|^2 \leq 2\mu^2 + \frac{\epsilon^2}{2} + 2\mu^2 + \frac{\epsilon^2}{2} - 4\mu^2.$$

This implies

$$\|k_j - k_i\|^2 \ \leq \epsilon^2.$$

It follows that

$$\|k_j - k_i\| \ \leq \ \epsilon$$

Thus, k_n forms a Cauchy Sequence. And, since K is a closed subset of Hilbert Space (which is complete), k_n is a converging sequence (due to completeness) with the limit point inside K (due to closedness). Thus there exists a $k_0 \in M$, such that

$$\|k_n - k_0\| \to 0 \text{ as } n \to \infty.$$

Since, $\|(x - k_n) - (x - k_0)\| = \|k_n - k_0\|$ we have

$$(x - k_n) \to (x - k_0) \text{ as } n \to \infty$$

From the continuity of norm, $\|(x - k_n)\|$ converges to $\|(x - k_0)\|$. Since, a converging sequence has a unique limit point, we have

$$\mu = \|x - k_0\|.$$

This proves the theorem ■

Separation of Disjoint Convex Sets

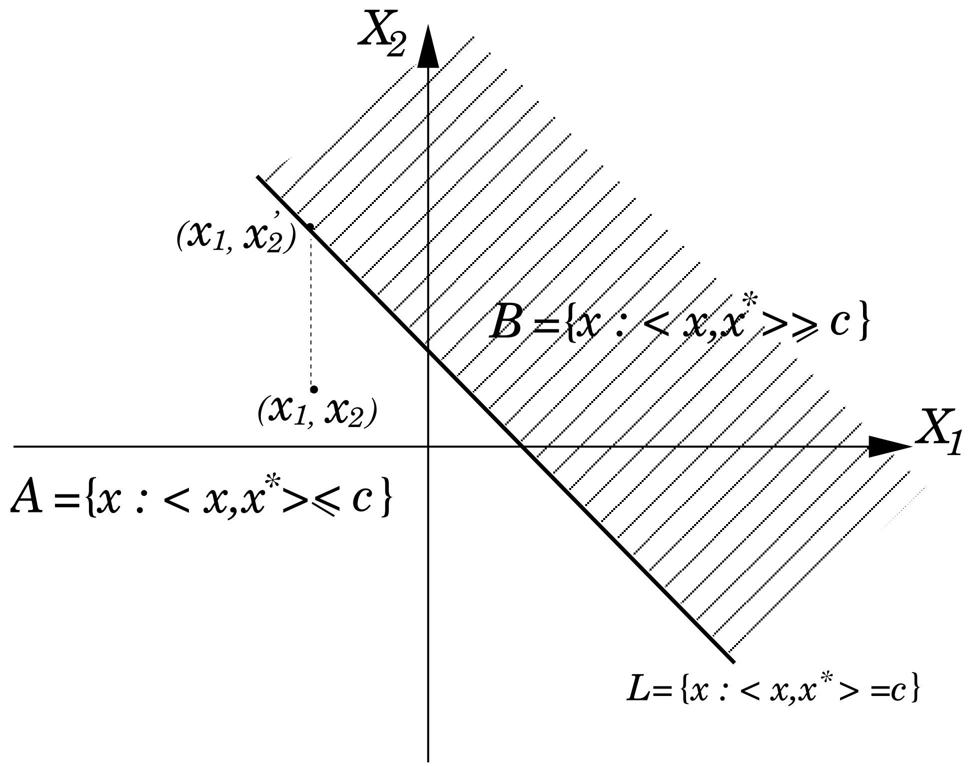

Figure 2: Separation of R^2 into half spaces by a line L.

Consider the vector space R^2. The equation of a line (see Figure 2) in R^2 is given by

$$m_1x_1 + m_2x_2 = c,$$

where m_1, m_2 and c are constants. The graph of a line is given by the set

$$L = \{(x_1, x_2)|m_1x_1 + m_2x_2 = c\},$$

which can be written as

$$L = \{x \in R^2| < x, x^* >= c\}, \tag{12}$$

where $x^* = (m_1, m_2)$. Note that if $m_2 = 0$ then we have a vertical line. We now generalize the concept of a line in R^2 to normed vector spaces. The line L defined earlier is a hyperplane in R^2.

Definition 27. *H is a hyperplane in a Hilbert Space X if and only if there exists a nonzero linear function $x^* : X \to R$ such that*

$$H := \{x :< x, x^* >= c\},$$

where $c \in R$.

For the purposes of the discussion below we will assume that c, m_1 and m_2 which describe the line L in Figure 12 are all nonnegative. The results for other cases will be similar. Consider the region A in Figure 12 which is the region "below" the line L. As illustrated earlier,

- $L = \{x :< x, x^* >= c\}$ where $x^* = (m_1, m_2)$.
- Consider any point $x = (x_1, x_2)$ in region A. Such a point lies "below" the line L. Thus if $x' = (x_1, x_2')$ denotes the point on the line L which has the same first coordinate as that of x then $x_2' \geq x_2$. As x' is on the line L it follows that $< x', x^* >= m_1x_1 + m_2x_2' = c$. As $m_2 \geq 0$ it follows that $< x, x^* >= m_1x_1 + m_2x_2 \leq m_1x_1 + m_2x_2' = c$. Thus we have shown that for every point x in the region A, $< x, x^* >\leq c$.
- In a similar manner it can be established that if $< x, x^* >\leq c$ then x lies "below" the line L, that is $x \in A$.

- Thus the region A is given by the set $\{x :< x, x^* > \leq c\}$ which is termed the negative half space of L.

- In an analogous manner it can be shown that the region B (which is the region “above” the line L) is described by $\{x :< x, x^* > \geq c\}$. This set is termed the positive half space of L.

Thus the line L separates R^2 into two halves; a positive and a negative half. We generalize the concept of half spaces for an arbitrary normed vector space.

Half spaces

Definition 28. [Half spaces] *Let $(X, ||\cdot||_X)$ be a normed linear space and let $x^* : X \to R$ be a bounded linear function on X. Let*

$$\begin{array}{rcl} S_1 & := & \{x \in X :< x, x^* > < c\}, \\ S_2 & := & \{x \in X :< x, x^* > \leq c\}, \\ S_3 & = & \{x \in X :< x, x^* > > c\}, \\ S_4 & := & \{x \in X :< x, x^* > \geq c\}. \end{array}$$

Then S_1 is an open negative half space, S_2 is a closed negative half space, S_3 is an open positive half space and S_4 is a closed positive half space.

Separation of convex sets in R^2.

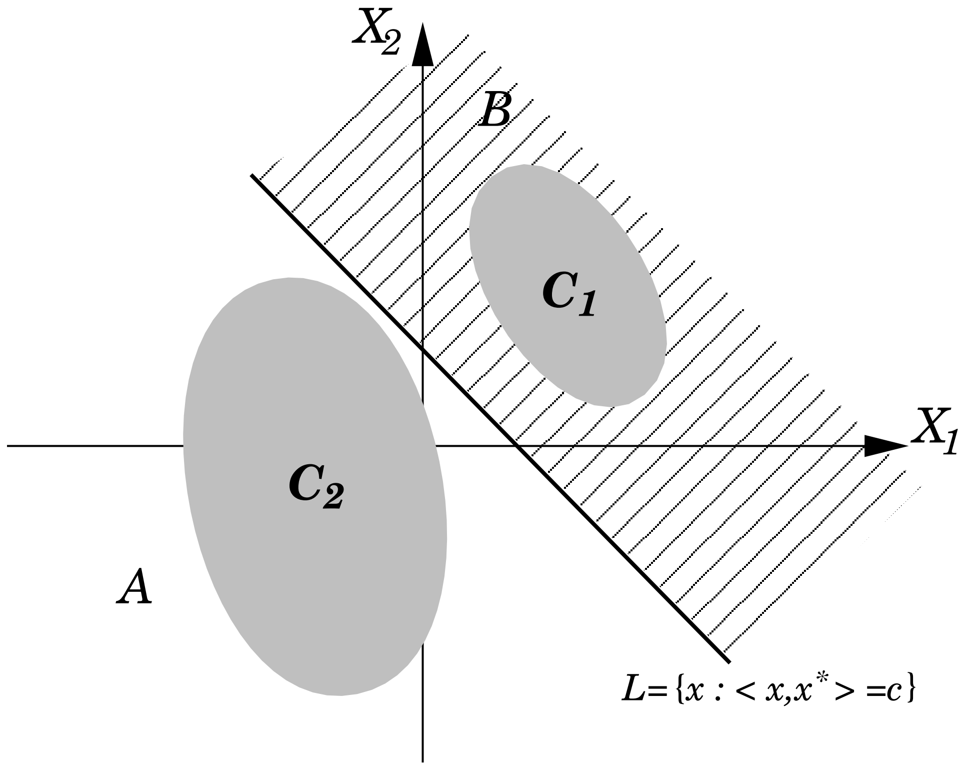

Figure 3: Separation of of convex sets in R^2.

It is intuitively clear that in R^2 if two convex sets C_1 and C_2 do not intersect then there exists a line in R^2 which separates the two sets (see Figure 3). In other words there exists x^* in $(R^2)^*$ and a constant c in R such that C_1 lies on the positive half space of the line $L = \{x| < x, x^* >= c\}$ and C_2 lies in the negative half space of L. That is

$$C_1 \subset \{x :< x, x^* > \geq c\},$$

and

$$C_2 \subset \{x :< x, x^* > \leq c\}.$$

The main focus of this section is to generalize this result to disjoint convex sets in a general normed vector space.

■

Separation of point and a closed convex set

Theorem 34. *Let K be a closed convex subset of a Hilbert space H. Let $x \in H$ be such that $x \notin K$. Then there exists a hyperplane that separates the point x from the set K. That is there exist an element $x^* \in H$ with $\|x^*\| = 1$ such that*

$$< k, x^* > \leq < x, x^* > \text{ for all } k \in K.$$

Proof: Let

$$\mu = \inf\{\|x - k\| : k \in K\}.$$

From 33 there exists $k_0 \in K$ such that $\|x - k_0\| = \mu$ and $< x - k_0, k - k_0 > \leq 0$ for all $k \in K$. Note that $\mu > 0$.

Indeed, if $\mu = 0$ then $\|x - k_0\| = \mu = 0$ and thus $x = k_0 \in K$ that contradicts the fact that $x \notin K$. Thus $\mu > 0$.

Let

$$x^* := \frac{x - k_0}{\|x - k_0\|}.$$

Note that for all $k \in K$

$$\begin{aligned}
< k, x^* > - < x, x^* > &= < k - x, x^* > \\
&= \frac{1}{\|x-k_0\|} < k - x, x - k_0 > \\
&= \frac{1}{\|x-k_0\|}(< k - k_0 + k_0 - x, x - k_0 >) \\
&= \frac{1}{\|x-k_0\|}(< x - k_0, k - k_0 > - \|x - k_0\|^2) \\
&< 0.
\end{aligned}$$

Thus

$$< k, x^* > \; < \; < x, x^* > \text{ for all } k \in K.$$

∎

A closed convex set characterized in terms of hyperplanes

Theorem 35. *If K is a closed convex set of a Hilbert space then K is equal to the intersection of all the closed half-spaces that contain it.*

Proof: Let the set Λ be a set such that $S_\lambda,\ \lambda \in \Lambda$ is a half-space that contains K and Λ characterizes all such sets. Let

$$S = \cap_{\lambda \in \Lambda} S_\lambda.$$

Suppose $k \in K$. Then $k \in S_\lambda$ for every $\lambda \in \Lambda$ as $S_\lambda \supset K$. Thus $k \in S$.

Suppose $h \notin K$. Then from Theorem 34 it follows that there exists a half space S_{λ_0} that contains K but does not contain h. Thus $h \notin S$. This proves the theorem.

∎

Aside: Banach-Alaoglu result

Lemma 15. *Let H be a separable Hilbert space. Consider*

$$B = \{x^* \in H : \|x^*\| \leq M\}.$$

Then for every sequence x^*k and $x \in H$ there exists a subsequence $x^*_{n_k}$

$$< x, x^*_{n_k} > \to < x, x^* > .$$

Proof: Let $x \in H$ be arbitrary. For this given x and any n

$$| < x, x^*_n > | \leq \|x\| \|x^* n\| \leq \|x\| M.$$

Thus the real numbers

$$< x, x^*_n > \in \{r \in R | |r| \leq \|x\| M\}.$$

Thus $r_n(x) =< x, x_n^* >$ is a sequence of real numbers that lies in a closed and bounded set. From a result in Real Analysis we can conclude that there exists a real number $r(x)$ and a subsequence of $r_{n_k}(x)$ such that $r_{n_k}(x) \to r(x) \ \in \ R.$

Suppose x and y are two elements in H. We will now show that $r(x+y) = r(x) + r(y)$ and $r(\alpha x) = \alpha r(x)$. Let n_k be a common subsequence such that

$$r_{n_k}(x) \to f(x),\ r_{n_k}(y) \to r(y) \text{ and } r_{n_k}(x+y) \to r(x+y).$$

Clearly

$$r_{n_k}(x) + r_{n_k}(y) \to r(x) + r(y).$$

Also, as

$$\begin{array}{rcl} r_{n_k}(x+y) & \to & r(x+y) \\ \Rightarrow < x+y, x_{n_k}^* > & \to & r(x+y) \\ \Rightarrow < x, x_{n_k}^* > + < y, x_{n_k}^* > & \to & r(x+y) \\ \Rightarrow r_{n_k}(x) + r_{n_k}(y) & \to & r(x+y) \end{array}$$

Thus we have

$$\begin{aligned} r_{n_k}(x) + r_{n_k}(y) &\rightarrow r(x+y) \text{ and} \\ r_{n_k}(x) + r_{n_k}(y) &\rightarrow r(x) + r(y). \end{aligned}$$

From the uniqueness of the limit point it follows that

$$r(x+y) = r(x) + r(y).$$

Similarly one can show that

$$r(\alpha x) = \alpha(x).$$

Thus r is a linear function on H. Also, for any x in H

$$|r(x)| = |\lim_{k\rightarrow\infty} < x, x^*_{n_k} > | \leq \lim_{k\rightarrow\infty} | < x, x^*_{n_k} > | \leq \|x\| \|x^*_{n_k} \leq \|x\| M.$$

Thus r is a linear function that has uniform bound M. Thus there exists an $x^* \in H$ such that

$$r(x) =< x, x^* >$$

that satisfies the property that for every $x \in H$ there exists a subsequence $x^*_{n_k}$ such that

$$< x, x^*_{n_k} > \rightarrow < x, x^* > .$$

∎

Separation of a point and the interior of a convex set

Theorem 36. *Let K be a convex subset of a Hilbert Space H. Let $x \in H$ and $x \notin int(H)$. Then there exists an element $x^* \in H$ such that $x^* \neq 0$ and*

$$< k, x^* > \leq < x, x^* > \text{ for all } k \in K.$$

Proof: We will prove this result only for separable Hilbert spaces. Let $cl(K)$ be the closure of the convex set K. If $x \in cl(K)$ then we obtain the result from Theorem 34. Suppose $x \in cl(K)$. As $x \notin int(K)$, $k \in bd(K)$. Thus there exists a sequence $x_n \in H$ with $x_n \notin cl(K)$ such that $\|x_n - x\| \to 0$. From Theorem 34 there exists $a_n \in H$ with $\|a_n\| = 1$ such that a_n separates x_n and $cl(K)$. That is

$$< k, a_n > \leq < x_n, a_n > \text{ for all } k \in cl(K).$$

Note that

$$a_n \in B^* := \{x^* \in H : \|x^*\| \leq 1\}.$$

From Lemma 15 there exists an $a \in B^*$ such that

$$< h, a_n > \rightarrow < h, a > \text{ for all } h \in H.$$

Note that from Theorem 1 we have

$$\lim_{n \rightarrow \infty} < x_n, a_n > = < x, a > .$$

Thus

$$\begin{array}{rlcl} & < k, a_n > & \leq & < x_n, a_n > \text{ for all } n, \text{ for all } k \in K \\ \Rightarrow & \lim_{n \rightarrow \infty} < k, a_n > & \leq & \lim_{n \rightarrow \infty} < x_n, a_n >, \quad \text{for all } k \\ \Rightarrow & < k, a > & \leq & < x, a >, \quad \text{for all } k \in K \end{array}$$

Note that $a \neq 0$ and thus a characterizes the hyperplane that separates $int(K)$ and $x \in bd(K)$.

Eidelheit Separation

Theorem 37. *Let K_1 and K_2 be convex subsets of a Hilbert space H with $int(K_1) \neq \{\}$ and $int(K_1) \cap K_2 = \{\}$. Then there is a hyperplane separating K_1 and K_2. That is there exists $x^* \in H$, $x^* \neq 0$ such that*

$$< k_1, x^* > \leq < k_2, x^* > \text{ for all } k_1 \in K_1 \text{ and } k_2 \in K_2.$$

Proof: Let $K = K_1 - K_2$. Then $int(K) \neq \{\}$ and $0 \in int(K)$. From Theorem 36 there exists a hyperplane characterized by $x^* \neq 0$ such that

$$< k, x^* > \leq 0 \text{ for all } k \in K.$$

That is

$$< k_1 - k_2, x^* > \leq 0 \text{ for all } k_1 \in K_1 \text{ and } k_2 \in K_2.$$

Thus

$$< k_1, x^* > \leq < k_2, x^* > \text{ for all } k_1 \in K_1 \text{ and } k_2 \in K_2.$$

This proves the theorem. ∎

Convex maps

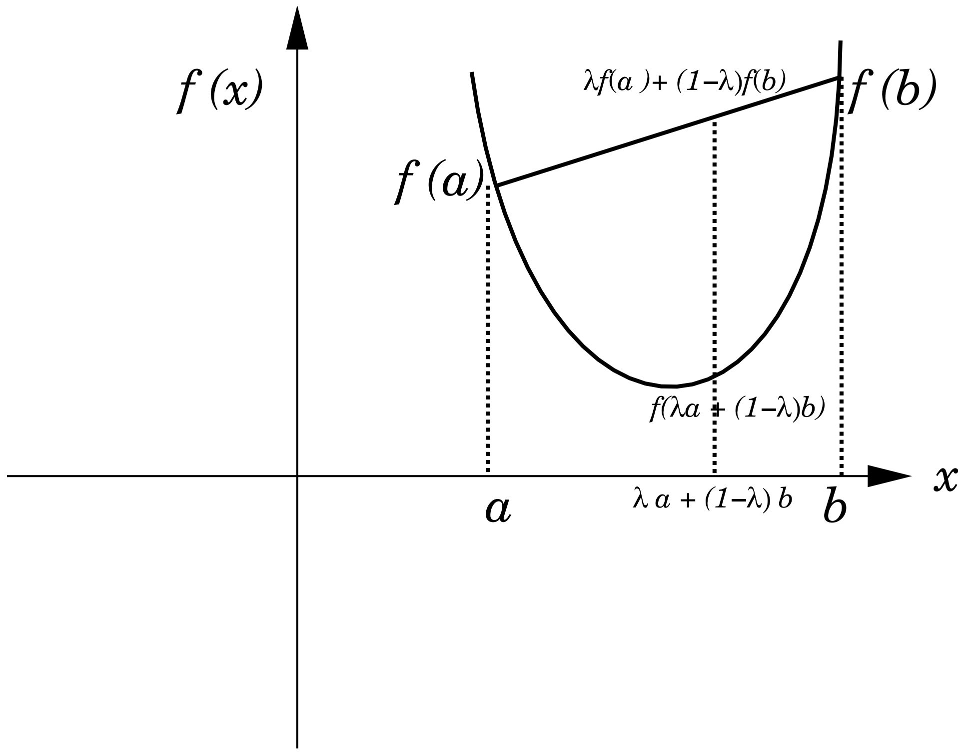

Figure 4: A convex function.

Definition 29. [Convex maps] *Let X be a vector space and Z be a vector space with positive cone P. A mapping,* $G : X \to Z$ *is convex if*

$G(tx + (1-t)y) \leq tG(x) + (1-t)G(y)$ *for all* x, y *in* X *and* t *with* $0 \leq t \leq 1$ *and is strictly convex if* $G(tx + (1-t)y) < tG(x) + (1-t)G(y)$ *for all* $x \neq y$ *in* X *and* t *with* $0 < t < 1$.

Epigraph

Definition 30. [Epigraph] *Let $f : \Omega \to R$ be a real valued function where Ω is a subset of a vector space X. The epigraph of f over Ω is a subset $[f, \Omega]$ of $R \times X$ defined by*

$$[f, \Omega] := \{(r, \omega) \in R \times X : x \in \Omega, f(x) \leq r\}.$$

Lemma 16. *Let $f : \Omega \to R$ be a real valued function where Ω is a convex subset of a vector space X. Then f is convex if and only if $[f, \Omega]$ is convex.*

Proof: Left to the reader. $\blacksquare$

Convex Optimization

The problem that is the subject of the rest of the chapter is the following problem.

$$\mu \quad = \quad \inf f(x) \quad \text{subject to} \quad x \in \Omega,$$

where $f : \Omega \to R$ is a convex function on a convex subset Ω of a vector space X. Such a problem is called a convex optimization problem.

Local minimum is global minimum

Lemma 17. *Let $f : (X, \|.\|_X) \to R$ be a convex function and let ω be a convex subset of X. If there exists a neighbourhood N in Ω of ω_0 where $\omega_0 \in \Omega$ such that for all $\omega \in N$, $f(\omega_0) \leq f(\omega)$ then $f(\omega_0) \leq f(\omega)$ for all ω in Ω (that is every local minimum is a global minimum).*

Proof: Let ω be any element of Ω. Let $0 \leq \lambda \leq 1$ be such that $x := \lambda\omega_0 + (1-\lambda)\omega$ be in N. Then $f(\omega_0) \leq f(x) \leq \lambda f(\omega_0) + (1-\lambda)f(\omega)$. This implies that $f(\omega_0) \leq f(\omega)$. As ω is an arbitrary element of Ω we have established the lemma. ■.

Uniqueness of the optimal solution

Lemma 18. *Let Ω be a convex subset of a Banach space X and $f : \Omega \to R$ be strictly convex. If there exists an $x_0 \in \Omega$ such that*

$$f(x_0) = \inf_{x \in \Omega} f(x),$$

(that is f achieves its minimum on Ω) then the minimizer is unique.

Proof: Let $m := \min_{x \in \Omega} f(x)$. Let $x_1, x_2 \in \Omega$ be such that $f(x_1) = f(x_2) = m$. Let $0 < \lambda < 1$. From convexity of Ω we have $\lambda x_1 + (1 - \lambda)x_2 \in \Omega$. From strict convexity of f we have that if $x_1 \neq x_2$ then $f(\lambda x_1 + (1 - \lambda)x_2) < \lambda f(x_1) + (1 - \lambda)f(x_2) = m$ which is a contradiction. Therefore $x_1 = x_2$. This proves the lemma. ∎

Varying the constraint level

Many convex optimization problems have the following structure

$$\omega(z) = \begin{array}{ll} & \inf f(x) \\ \text{subject to} & \\ & x \in \Omega \\ & g(x) \leq z, \end{array} \qquad (13)$$

where $f : \Omega \to R$, $g : X \to Z$ are convex maps with Ω a convex subset of the vector space X and Z a normed vector space with a positive cone P. The condition $g(x) \leq z$ is to be interpreted with respect to the positive cone P of the vector space Z.

Lemma 19. *The function ω is convex.*

Proof: Let z_1 and z_2 be elements in Z and let $0 \le \lambda \le 1$ be any constant. Then

$$
\begin{aligned}
\omega(\lambda z_1 + (1-\lambda) z_2) &= \inf\{f(x) : x \in \Omega, g(x) \le \lambda z_1 + (1-\lambda) z_2\} \\
&= \inf\{f(x) : x = \lambda x_1 + (1-\lambda) x_2, x_1 \in \Omega, x_2 \in \Omega, \\
&\qquad g(x) \le \lambda z_1 + (1-\lambda) z_2\} \\
&\le \inf\{\lambda f(x_1) + (1-\lambda) f(x_2), x_1 \in \Omega, x_2 \in \Omega, \\
&\qquad g(x) \le \lambda z_1 + (1-\lambda) z_2\} \\
&\le \inf\{\lambda f(x_1) + (1-\lambda) f(x_2), x_1 \in \Omega, x_2 \in \Omega, \\
&\qquad g(x_1) \le z_1, g(x_2) \le z_2\} \\
&= \lambda \omega(z_1) + (1-\lambda)\omega(z_2).
\end{aligned}
$$

The second equality is true because for any given λ with $0 \le \lambda \le 1$ the set $\Omega = \{x : x = \lambda x_1 + (1-\lambda) x_2, x_1 \in \Omega, x_2 \in \Omega\}$. The first inequality is true because f is a convex map. The second inequality is true because the set $\{(x_1, x_2) \in \Omega \times \Omega : g(\lambda x_1 + (1-\lambda) x_2) \le \lambda z_1 + (1-\lambda) z_2\} \supset \{(x_1, x_2) \in \Omega \times \Omega : g(x_1) \le z_1, g(x_2) \le z_2\}$, which follows from the convexity of g. This proves the lemma. ∎

Lemma 20. *Let* z_1 *and* z_2 *be elements in* Z *such that* $z_1 \leq z_2$ *with respect to the convex cone* P*. Then* $\omega(z_2) \leq \omega(z_1)$.

Proof: Follows immediately from the relation
$\{x \in \Omega : g(x) \leq z_2\} \supset \{x \in \Omega : g(x) \leq z_1\}$, if $z_1 \leq z_2$. ∎

Kuhn-Tucker Theorem

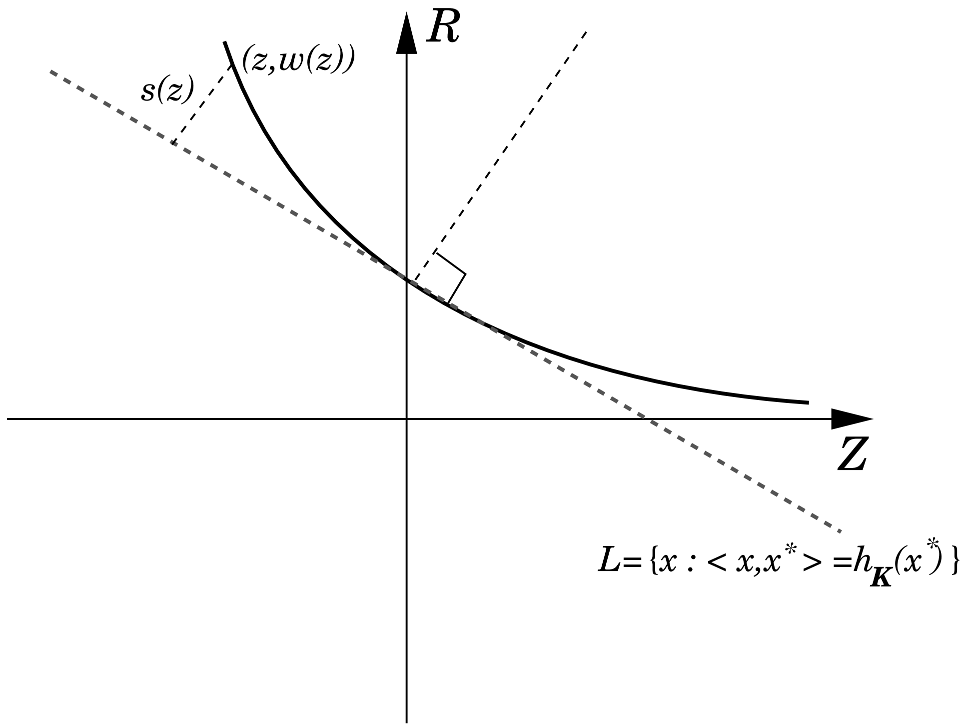

Figure 5: Illutstration of $\omega(z)$.

Consider the convex optimization problem

$$\begin{aligned} \omega(z) \quad = \quad & \inf f(x) \\ \text{subject to} \quad & \\ & x \in \Omega \\ & g(x) \leq z. \end{aligned}$$

We will obtain information about $\omega(0)$ by analyzing $\omega(z)$. We have shown that $\omega(z)$ is a decreasing function of z (see Lemma 20) and that it is a convex function (see Lemma 19). It can be visualized as illustrated in Figure 5. As $\omega(z)$ is a decreasing function it is evident that the tangent to the curve at $(0, \omega(0))$ has a negative slope (see Figure 5). Thus the tangent can be characterized by a line L with the equation:

$$\omega(z) + < z, z^* >= c,$$

where $z^* \geq 0$. Also, note that if we change the coordinates such that L becomes the horizontal axis and its perpendicular the vertical axis with the origin at $(0, \omega(0))$ (see Figure 5) then the function $\omega(z)$ achieves its minimum at the new origin. In the new cordinate system the vertical cordinate of the curve $\omega(z)$ is given by the distance of $(z, \omega(z))$ from the line L. This distance is given by

$$s(z) = \frac{\omega(z) + < z, z^* > -c}{\|(1, z^*)\|}.$$

Thus $s(z)$ achieves its minimum at $z = 0$. This implies that

$$\begin{array}{rcl}
\omega(0) & = & \min_{z \in Z} \{\omega(z) + < z, z^* >\} \\
& = & \min_{z \in Z} \{\inf\{f(x) : x \in \Omega, g(x) \leq z\} + < z, z^* >\} \\
& = & \inf\{f(x) + < z, z^* >: x \in \Omega, z \in Z, g(x) \leq z\} \\
& \geq & \inf\{f(x) + < g(x), z^* >: x \in \Omega, z \in Z, g(x) \leq z\} \\
& \geq & \inf\{f(x) + < g(x), z^* >: x \in \Omega\}.
\end{array}$$

The first inequality is true because $z^* \geq 0$ and $g(x) \leq z$. The second inequality is true because the $\{x \in \Omega : z \in Z, g(x) \leq z\} \subset \{x \in \Omega\}$. It is also true that

$$\begin{aligned} \omega(0) &= \inf\{f(x) + < z, z^* >: x \in \Omega, z \in Z, g(x) \leq z\} \\ &\leq \inf\{f(x) + < g(x), z^* >: x \in \Omega\}, \end{aligned}$$

because $g(x) \leq g(x)$ is true for every $x \in \Omega$. Thus we have

$$\omega(0) = \inf\{f(x) + < z, z^* >: x \in \Omega\}.$$

Note that the above equation states that a constrained optimization problem given by the problem statement of $\omega(0)$ can be converted to an unconstrained optimization problem as given by the right hand side of the above equation. We make these arguments more precise in the rest of this subsection.

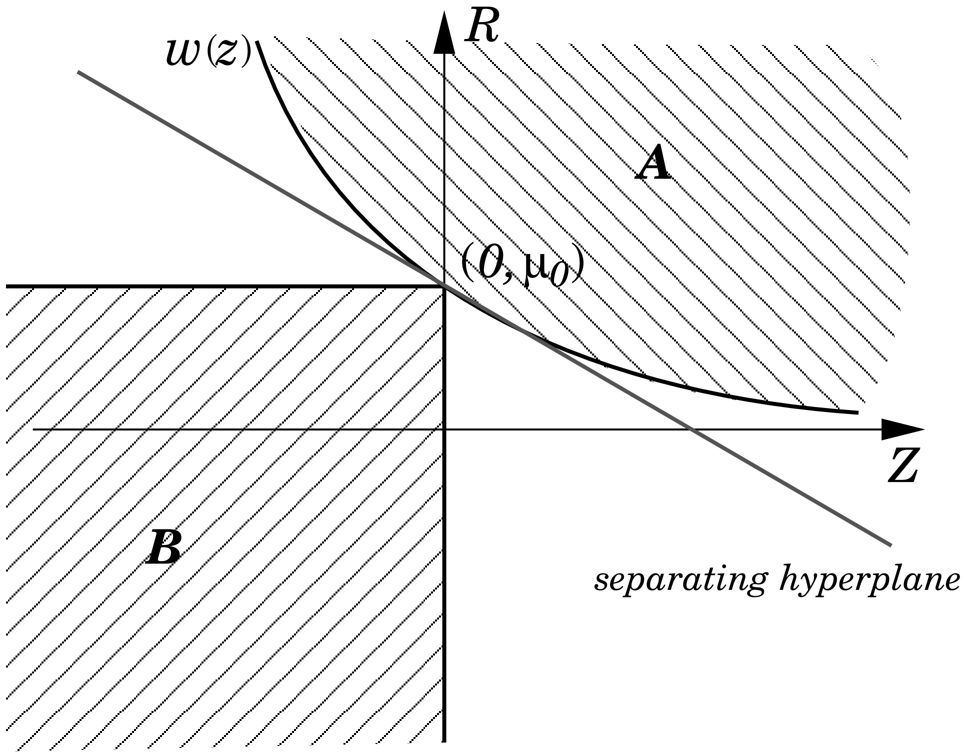

Figure 6: Figure for Lemma 21.

Lemma 21. *Let $(X, ||\cdot||_X)$, and $(Z, ||\cdot||_Z)$, be normed vector spaces with Ω a convex subset of X. Let P be a positive convex cone defined in Z. Let Z^* denote the dual space of Z with the postive cone $P^{\oplus}$ associated with P. Let*

$f : \Omega \to R$ be a real valued convex functional and $g : X \to Z$ be a convex mapping. Define

$$\mu_0 := \inf\{f(x) : \ g(x) \leq 0, \ x \in \Omega\}. \tag{14}$$

Suppose there exists $x_1 \in \Omega$ such that $g(x_1) < 0$ and and suppose μ_0 is finite. Then, there exist $z_0^* \geq 0$ such that

$$\mu_0 = \inf\{f(x) + < g(x), z_0^* >: x \in \Omega\}. \tag{15}$$

Furthermore, if there exists x_0 such that $g(x_0) \leq 0$ and $\mu_0 = f(x_0)$ then

$$< g(x_0), z_0^* >= 0 \tag{16}$$

Proof: We will say that an element x in Ω is feasible if $g(x) \leq 0$. Define A, (see Figure 6) a subset of $Z \times R$ by

$$A := \{(z, r) : \text{ there exists } x \in \Omega \text{ such that } g(x) \leq z \text{ and } f(x) \leq r\},$$

and B (see Figure 6) another subset of $Z \times R$ by

$$B := -P \times (-\infty, \mu_0] := \{(z, r) : -z \in P \text{ and } r \leq \mu_0\}.$$

We will assume that the norm on $Z \times R$ is the product norm induced by the norms on Z and R. Note that in this norm $int(B) \neq \{\}$ (let $p_0 \in int(-P)$; then $(p_0, \mu_0 - 1) \in int(B)$). We will show that $int(B) \cap A = \{\}$.

Suppose $(z, r) \in int(B) \cap A$. Then there exists x in Ω such that $f(x) \leq r$ and $g(x) \leq z$. Also $z \in -P$ and $r < \mu_0$. Therefore, $f(x) \leq r < \mu_0$ and $g(x) \leq z \leq 0$. This implies that x is feasible and $f(x)$ is strictly less than μ_0 which contradicts the definition of μ_0. Therefore, $int(B) \cap A = \{\}$.

Applying Eidelheit's separation result (see Corollary **??**) to A and B (note that A and B are convex) we know that there exists a nonzero element $(z^*, s) \in (Z \times R)^* = Z^* \times R$ (see Theorem **??**) and $k \in R$ such that

$$< z, z^* > +sr \geq k \text{ for all } (z, r) \in A \text{ and} \tag{17}$$

$$< z, z^* > + sr \leq k \text{ for all } (z, r) \in B. \tag{18}$$

We will now show that $s \geq 0$. As $(0, r)$ for $r \leq \mu_0$ is in B it follows from inequality (18) that $sr \leq k$ for all $r \leq \mu_0$. This implies that $s \geq 0$ (otherwise by letting $r \to -\infty$ we see that $k = \infty$ which is not possible because inequality (17) holds).

We will now show that $s > 0$. Suppose that $s = 0$. Then from inequality (17) we have

$$< g(x_1), z^* > \geq k, \tag{19}$$

because $(g(x_1), f(x_1))$ belongs to A. Also, from inequality (18) we have that

$$< z, z^* > \leq k, \tag{20}$$

for all $z \in -P$. In particular as $0 \in -P$ we have $k \geq 0$. Suppose for some $z \in -P$, $< z, z^* > > 0$. Then we have $< \alpha z, z^* > = \alpha < z, z^* > \to \infty$ as

$\alpha \to \infty$. However as P is a cone and $\alpha \geq 0$, $\alpha z \in -P$ if $z \in -P$. Therefore $< \alpha z, z^* > \leq\ k < \infty$ if $z \in -P$. Thus we have a contradiction and therefore

$$< z, z^* > \leq 0 \text{ for all } z \in -P \text{ and } k \geq 0. \tag{21}$$

As $-g(x_1) \in int(P)$ we have that there exists an $\epsilon > 0$ in R such that $||z||_Z \leq \epsilon$ implies that $-g(x_1) + z \in P$. Therefore, from (21) we have that $< g(x_1) - z, z^* > \leq 0$ if $||z||_Z \leq \epsilon$ which implies that $< g(x_1), z^* > \leq < z, z^* >$ if $||z||_Z \leq \epsilon$. From inequality (19) we have $0 \leq k \leq < g(x_1), z^* > \leq < z, z^* >$ if $||z||_Z \leq \epsilon$. This implies that for any $z \in Z$, $< z, z^* > \geq 0$. For any nonzero $z \in Z$,

$$|| \frac{\epsilon z}{||z||_Z} ||_Z \leq \epsilon$$

and therefore $< \frac{\epsilon z}{||z||_Z}, z^* > \geq 0$. This implies that for any $z \in Z$, $< z, z^* > \geq\ 0$. As Z is a vector space (which implies $- < z, z^* > \geq\ 0)$ it follows that $< z, z^* > = 0$ for all $z \in Z$. Thus $z^* = 0$. This contradicts $(z, s) \neq (0, 0)$ and therefore, $s > 0$.

Let $z_0^* = \frac{z^*}{s}$. Dividing inequality (17) by s we have

$$< z, z_0^* > +r \geq \frac{k}{s} \text{ for all } (z, r) \in A \text{ and} \tag{22}$$

dividing inequality (18) by s we have

$$< z, z_0^* > +r \leq \frac{k}{s} \text{ for all } (z, r) \in B. \tag{23}$$

In particular, as $(z, \mu_0) \in B$ for all $z \in -P$ it follows from inequality (23) that

$$< z, z_0^* > \leq \frac{k}{s} - \mu_0 \text{ for all } z \in -P.$$

This implies that $< z, z_0^* > \leq\ 0$ for all $z \in -P$. Indeed, if for some $z_1 \in -P, < z_1, z_0^* > >\ 0$ then $< \alpha z_1, z^* > \to \infty$ as $\alpha \to \infty$ which contradicts the fact that $< \alpha z_1, z^* >$ is bounded above by $\frac{k}{s} - \mu_0$. Thus we conclude that $z_0^* \in P^{\oplus}$.

Also, as $(g(x), f(x))$ for $x \in \Omega$ is in A it follows from (22) that

$$< g(x), z_0^* > + f(x) \geq \frac{k}{s} \text{ for all } x \in \Omega \text{ and} \tag{24}$$

as $(0, \mu_0) \in B$ it folllows from (23) that

$$\mu_0 \leq \frac{k}{s} \text{ for all } (z, r) \in B. \tag{25}$$

From inequalities (24) and (25) we conclude that

$$\inf\{< g(x), z_0^* > + f(x) : x \in \Omega\} \geq \mu_0. \tag{26}$$

Suppose $x \in \Omega$ and $g(x) \leq 0$ (i.e. x is feasible), then

$$f(x) + < g(x), z_0^* > \leq f(x), \tag{27}$$

because $z_0^* \in P^{\oplus}$. Therefore, we have

$$\begin{aligned}
\inf\{f(x) + < g(x), z_0^* >: x \in \Omega\} &\leq \inf\{f(x) + < g(x), z_0^* > \\
&\qquad : x \in \Omega, g(x) \leq 0\} \\
&\leq \inf\{f(x) : x \in \Omega, g(x) \leq 0\} = \mu_0.
\end{aligned}$$

The first inequality is true because $\Omega \supset \{x \in \Omega, g(x) \leq 0\}$ and the second inequality follows from (27).

It follows from inequality (26) that

$$\mu_0 = \inf\{f(x) + < g(x), z_0^* >: x \in \Omega\}. \tag{28}$$

Let x_0 be such that $x_0 \in \Omega$ and $g(x_0) \leq 0$ and $f(x_0) = \mu_0$. Then

$$f(x_0) = \mu_0 \leq f(x_0) + < g(x_0), z_0^* > \leq \ f(x_0) = \mu_0.$$

The first inequality follows from equation (28) and the second inequality is true because $z_0^* \in P^{\oplus}$ and $g(x_0) \leq 0$. This proves that $< g(x_0), z_0^* >= 0$. ■

Theorem 38. *Let X be a Banach space, Ω be a convex subset of X, Y be a finite dimensional normed space, Z be a normed space with positive cone P. Let Z^* denote the dual space of Z with a positive cone $P^{\oplus}$. Let $f : \Omega \to R$ be a real valued convex functional, $g : X \to Z$ be a convex mapping, $H : X \to Y$ be an affine linear map and $0 \in int(\{y \in Y : \ H(x) = y \text{ for some } x \in \Omega\})$. Define*

$$\mu_0 := \inf\{f(x) : \ g(x) \leq 0, \ H(x) = 0, \ x \in \Omega\}. \tag{29}$$

Suppose there exists $x_1 \in \Omega$ such that $g(x_1) < 0$ and $H(x_1) = 0$ and suppose

μ_0 *is finite. Then, there exist* $z_0^* \geq 0$ and y_0^* *such that*

$$\mu_0 = \inf\{f(x) + < g(x), z_0^* > + < H(x), y_0^* >: x \in \Omega\}. \tag{30}$$

Proof: Let

$$\Omega_1 := \{x : \ x \in \Omega, \ H(x) = 0\}.$$

Applying Lemma 21 to Ω_1 we know that there exists $z_0^* \in P^{\oplus}$ such that

$$\mu_0 = \inf\{f(x) + < g(x), z_0^* >: x \in \Omega_1\}. \tag{31}$$

Consider the convex subset,

$$H(\Omega) := \{y \in Y : \ y = H(x) \text{ for some } x \in \Omega\}$$

of Y. For $y \in H(\Omega)$ define

$$k(y) := \inf\{f(x) + < g(x), z_0^* >: \ x \in \Omega, \ H(x) = y\}.$$

We now show that k is convex. Suppose $y, y' \in H(\Omega)$ and x, x' are such that $H(x) = y$ and $H(x') = y'$. Suppose, $0 < \lambda < 1$. We have, $\lambda(f(x) + < g(x), z_0^* >) + (1-\lambda)(f(x') + < g(x'), z_0^* >) \geq f(\lambda x + (1-\lambda)x') + < g(\lambda x + (1-\lambda)x'), z_0^* > \geq k(\lambda y + (1-\lambda)y')$. (the first inequality follows from the convexity of f and g. The second inequality is true because $H(\lambda x + (1-\lambda)x') = \lambda y + (1-\lambda)y'$.) Taking infimum on the left hand side we obtain $\lambda k(y) + (1-\lambda)k(y') \geq k(\lambda y + (1-\lambda)y')$. This proves that k is a convex function.

We now show that $k : H(\Omega) \to R$ (i.e. we show that $k(y) > -\infty$ for all $y \in H(\Omega)$). As, $0 \in int[H(\Omega)]$ we know that there exists an $\epsilon > 0$ such that $\text{if } ||y|| \leq \epsilon \text{ then } y \in H(\Omega)$. Take any $y \in H(\Omega)$ such that $y \neq 0$. Choose λ, y' such that

$$\lambda = \frac{\epsilon}{2||y||} \text{ and } y' = -\lambda y.$$

This implies that $y' \in H(\Omega)$. Let, $\beta = \dfrac{\lambda}{\lambda + 1}$. We have

$$(1-\beta)y' + \beta y = 0.$$

Therefore, from convexity of the function k we have

$$\beta k(y) + (1-\beta)k(y') \geq k(0) = \mu_0.$$

Note that $\mu_0 > -\infty$ by assumption. Therefore, $k(y) > -\infty$. Note, that for all $y \in H(\Omega),\ k(y) < \infty$. This proves that k is a real valued function.

Let $[k, H(\Omega)]$ be defined as given below

$$[k, H(\Omega)] := \{(r, y) \in R \times Y : y \in H(\Omega),\ k(y) \leq r\}.$$

We first show that $[k, H(\Omega)]$ has nonempty interior. As, k is a real valued convex function on the finite-dimensional convex set $H[\Omega]$ and $0 \in int[H(\Omega)]$ we have from from Lemma **??** that k is continuous at 0. Let $r_0 = k(0) + 2$ and choose ϵ' such that $0 < \epsilon' < 1$. As, k is continuous at 0 we know that there exists $\delta > 0$ such that $y \in H(\Omega) \text{ and } ||y|| \leq \delta \text{ implies that}$

$$|k(y) - k(0)| < \epsilon'.$$

This means that if $y \in H(\Omega)$ and $||y|| \leq \delta$ then

$$k(y) < k(0) + \epsilon' < k(0) + 1 < r_0 - \frac{1}{2}.$$

Therefore, for all $y \in H(\Omega)$ with $||y|| \leq \delta$ we have $k(y) < r_0 - \frac{1}{2}$. This implies that for all $(r, y) \in R \times Y$ such that $|r - r_0| < \frac{1}{4}, y \in H(\Omega)$ and $||y|| \leq \delta$ we have $k(y) < r$. This proves that $(r_0, 0) \in int([k, H(\Omega)])$.

It is clear that $(k(0), 0) \in R \times Y$ is not in the interior of $[k, H(\Omega)]$. Using, Corollary **??** we know that there exists $(s, y^*) \neq (0, 0) \in R \times Y^*$ such that for all $(r, y) \in [k, H(\Omega)]$ the following is true

$$< y, y^* > +rs \geq < 0, y^* > +k(0)s = s\mu_0. \tag{32}$$

In particular, $rs \geq s\mu_0$ for all $r \geq \mu_0$ (note that $(r, 0) \in [k, H(\Omega)]$ for all $r \geq \mu_0$). This means that $s \geq 0$.

Suppose, $s = 0$. We have from (32) that $< y, y^* > \geq 0$ for all $y \in H(\Omega)$. As, $0 \in int[H(\Omega)]$ it follows that there exists an $\epsilon \in R$ such that $||y|| \leq \epsilon$ implies that $< y, y^* > \geq 0$ and $< -y, y^* > \geq 0$. This implies that if $||y|| \leq \epsilon$ then $< y, y^* > = 0$. But, then for any $y \in Y$ one can choose a positive constant α such that $||\alpha y|| \leq \epsilon$ and therefore $< \alpha y, y^* > = 0$. This implies that $(s, y^*) = (0, 0)$ which is not possible. Therefore, we conclude that $s > 0$.

Let $y_0^* = y^*/s$. From (32) we have,

$$< y, y_0^* > + r \geq \mu_0, \quad \text{for all } (r, y) \in [k, H(\Omega)]. \tag{33}$$

This implies that $\text{for all } y \in H(\Omega)$,

$$< y, y_0^* > + k(y) \geq \mu_0, \tag{34}$$

(This is because $(k(y), y) \in [k, H(\Omega)]$). Therefore, $\text{for all } x \in \Omega$,

$$< H(x), y_0^* > + f(x) + < g(x), z_0^* > \geq \mu_0, \tag{35}$$

which implies that

$$\inf\{f(x)+ < g(x), z_0^* > + < H(x), y_0^* >: \; x \in \Omega\} \geq \mu_0. \tag{36}$$

But if $x \in \Omega$ is such that $H(x) = 0$ then

$$\begin{aligned} f(x)+ < g(x), z_0^* > &= f(x)+ < g(x), z_0^* > + < H(x), y_0^* > \\ &\geq \inf\{f(x)+ < g(x), z_0^* > + < H(x), y_0^* >: \; x \in \Omega\} \\ &\geq \mu_0. \end{aligned}$$

Taking infimum on the left hand side of the above inequality over all $x \in \Omega$ which satisfy $H(x) = 0$ (that is infimum over all $x \in \Omega_1$) we have,

$$\mu_0 = \inf\{f(x)+ < g(x), z_0^* > + < H(x), y_0^* >: \; x \in \Omega\}. \tag{37}$$

This proves the lemma. ■

Primal Problem

Consider the problem

$$\mu_0 := \inf\{f(x) : \ g(x) \le 0,\ H(x) = 0,\ x \in \Omega\}. \tag{38}$$

The above problem is often called the as the *Primal* problem.

Lagrangian

Consider the primal problem

$$\mu_0 := \inf\{f(x) : \; g(x) \leq 0, \; H(x) = 0, \; x \in \Omega\}.$$

Associated with the above problem one can define the *Lagrangian*

$$L(x, z^*, y) := f(x) + < g(x), z^* > + < H(x), y > \tag{39}$$

where z^* is the dual variable associated with the constraint $g(x) \leq 0$ and y is the dual variable associated with the constraint $H(x) = 0$ for the primal problem. x is the primal variable.

The saddle point characterization: Existence in the Convex Case

Consider the problem setup in Theorem 38 and assume that x_0 is such that

$$x_0 = arg[\inf_{x \in \Omega} \{f(x) : G(x) \leq 0, H(x) = 0\}.]$$

Then there exists a $z_0^* \geq 0$ such that

$$L(x_0, z^*, y) \leq L(x_0, z_0^*, y) \leq L(x, z_0^*, y) \text{ for all } x \in \Omega, z^* \geq 0 \text{ and } y \in Y. \quad (40)$$

Proof: Proof: Let $z_0^* \geq 0, y$ be the dual variables as obtained in Theorem 38.

Then from Theorem 42 it follows that

$$\begin{aligned}
< g(x_0), z_0^* > &= 0 \\
g(x_0) &\leq 0 \\
H(x_0) &= 0 \\
\inf_{x\in\Omega}\{f(x) : G(x) \leq 0, H(x) = 0\} &= \inf_{x\in\Omega} L(x, z_0^*, y)
\end{aligned}$$

Thus

$$\begin{aligned}
L(x, z_0^*, y) \geq \inf_{x\in\Omega} L(x, z_0^*, y) &= \inf_{x\in\Omega}\{f(x) : G(x) \leq 0, H(x) = 0\} \\
&= f(x_0) \\
&= f(x_0) + < g(x_0), z_0^* > + < H(x_0), y > \\
&= L(x_0, z_0^*, y).
\end{aligned}$$

Also note that

$$
\begin{aligned}
L(x_0, z^*, y) - L(x_0, z_0^*, y) &= f(x_0) + < g(x_0), z^* > + < H(x_0), y) \\
&\quad -[f(x_0) + < g(x_0), z_0^* > + < H(x_0), y) \\
&= < g(x_0), z^* > \\
&\leq 0 \text{ for all } z^* \geq 0 \text{ as } g(x_0) \leq 0.
\end{aligned}
$$

$\blacksquare$

The saddle point characterization: Sufficiency of optimality

We need the following lemma to establish the optimality from the existence of a saddle point.

Lemma 22. *Suppose Z is a Hilbert space with a cone P defined such that P is closed. Let $P^{\oplus}$ be the dual cone that is*

$$P^{\oplus} := \{z^* \in Z | < z, z^* > \geq 0 \text{ for all } z \in P\}.$$

Suppose $z \in Z$ is such that

$$< z, z^* > \geq 0 \text{ for all } z^* \geq 0.$$

Then

$$z \geq 0.$$

Proof: Suppose $z \notin P$. Then as P is closed there exists a hyperplane that strictly separates z and P (see Theorem 34). That is there exists $z^* \neq 0$ and k

such that

$$< z, z^* >< k \leq < p, z^* > \text{ for all } p \in P.$$

This implies that

$$< z, z^* >< k \leq \inf_{p \in P} < p, z^* > \leq 0.$$

Note that it has to be true that $z^* \geq 0$. Otherwise suppose $p_1 \in P$ be such that $< p_1, z^* >= a < 0$. Then note that $< \alpha p_1, z^* >= -\alpha a \to -\infty$ as $\alpha \to \infty$. Note that $\alpha p_1 \in P$ for all $\alpha \geq 0$ and this will contradict that $< p, z^* >$ is lower bounded by k.

Thus we have found a $z^* \geq 0$ such that

$$< z, z^* >< 0$$

that contradicts the hypothesis on z. Thus $z \geq 0$.

∎

Theorem 39. *Let* $f : \Omega \to R$ *where* Ω *is a subset of a vector space* X. *Let* $g : \Omega \to Z$ *where* Z *is a normed vector space with positive cone* P *defined that is nonempty and closed. Suppose there exists a* $x_0 \in \Omega$ *and* $z_0^* \in P^{\oplus}$ *such that the Lagrangian possesses a saddle point at* x_0, z_0^*, y_0^* *that is*

$$L(x_0, z^*, y^*) \leq L(x_0, z_0^*, y_0^*) \leq L(x, z_0^*, y_0^*) \text{ for all } x \in \Omega,\ z^* \geq 0 \text{ and } y^* \in Y.$$

Then x_0 *is the optimal solution to the primal*

$$\mu_0 = \inf_{x \in \Omega} \{f(x) : g(x) \leq 0, H(x) = 0\}.$$

Proof: Note that

$$L(x_0, z^*, y^*) - L(x_0, z_0^*, y_0^*) \quad \leq \quad 0 \text{ for all } z^* \geq 0, \text{ and } y^* \in Y$$

Thus

$$< g(x_0), z^* - z_0^* > + < H(x_0), y^* - y_0^* > \leq \text{ for all } z^* \geq 0, \text{ and } y^* \in Y \quad (41)$$

By setting $y^* = y_0^*$ above we have

$$\begin{aligned} & < g(x_0), z^* - z_0^* > && \leq \quad \text{for all } z^* \geq 0 \\ \Rightarrow \quad & < g(x_0), (z_1^* + z_0^*) - z_0^* > && \leq \quad 0 \text{ for all } z_1^* \geq 0 \\ \Rightarrow \quad & < g(x_0), z_1^* > && \leq \quad 0 \text{ for all } z_1^* \geq 0 \\ \Rightarrow \quad & < -g(x_0), z^* > && \geq \quad 0 \text{ for all } z^* \geq 0. \end{aligned}$$

From Lemma 22 as the positive cone P is closed it follows that $-g(x_0) \geq 0$ that is $g(x_0) \leq 0$. Also be setting $z^* = 0$ in

$$< g(x_0), z^* - z_0^* > \quad \leq \quad \text{for all } z^* \geq 0$$

we have

$$< g(x_0), z_0^* > \geq 0.$$

As $g(x_0) \leq 0$ and $z_0^* \geq 0$ it follows that

$$< g(x_0), z_0^* > = 0.$$

By setting $z^* = z_0^*$ and $y^* = y_1^* + y_0^*$ in 41 we have

$$\begin{array}{rlcl} & < H(x_0), y_1^* > & \leq & 0 \text{ for all } y_1^* \in Y \\ \Rightarrow & < H(x_0), -y_1^* > & \leq & 0 \text{ for all } y_1^* \in Y \text{ as } -y_1^* \in Y \text{ if } y_1^* \in Y \\ \Rightarrow & < H(x_0), y_1^* > & = & 0 \text{ for all } y_1^* \in Y \\ \Rightarrow & < H(x_0), H(x_0) > & = & 0 \text{ as } H(x_0) \in Y. \end{array}$$

Thus we have shown that
$g(x_0) \leq 0,\ H(x_0) = 0$ and $< g(x_0), z_0^* > + < H(x_0), y_0^* > = 0$. Note that if

$x_1 \in \Omega$ is a feasible solution to the problem

$$\inf\{f(x) : x \in \Omega, g(x) \leq 0,\ H(x) = 0\}$$

then $x_1 \in \Omega,\ g(x_1) \leq 0$ and $H(x_1) = 0$. We also have

$$\begin{array}{rrcl}
 & L(x_1, z_0^*, y_0^*) - L(x_0, z_0^*, y_0^*) & \geq & 0 \\
\Rightarrow & f(x_1) + < g(x_1), z_0^*) + < H(x_1), y_0^* > & & \\
 & -[f(x_0) + < g(x_0), z_0^*) + < H(x_0), y_0^* >] & \geq & 0 \\
\Rightarrow & f(x_1) + < g(x_1), z_0^*) - f(x_0) & \geq & f(x_1) - f(x_0) \geq 0 \\
 & & & \text{ as } g(x_1) \leq \text{ and } z_0^* \geq 0 \\
\Rightarrow & f(x_1) & \geq & f(x_0)
\end{array}$$

Thus

$$x_0 = arg[\inf\{f(x) : x \in \Omega, g(x) \leq 0,\ H(x) = 0\}].$$

∎

Dual Interpretation

Consider the primal problem:

$$\omega(z) = \inf_{x \in \Omega} \{f(x) : g(x) \leq z\}.$$

We have shown that

- $\omega : Z \to R$ is a decreasing function of the variable $z \in Z$ that has a cone P defined. Thus $\omega(z_2) \leq \omega(z_1)$ if $z_2 \geq z_1$. Thus it can be assumed that the "slope" of the curve $\omega(z)$ is negative at any point z.

- $\omega(z)$ is a convex function of z if f and g are convex functions.

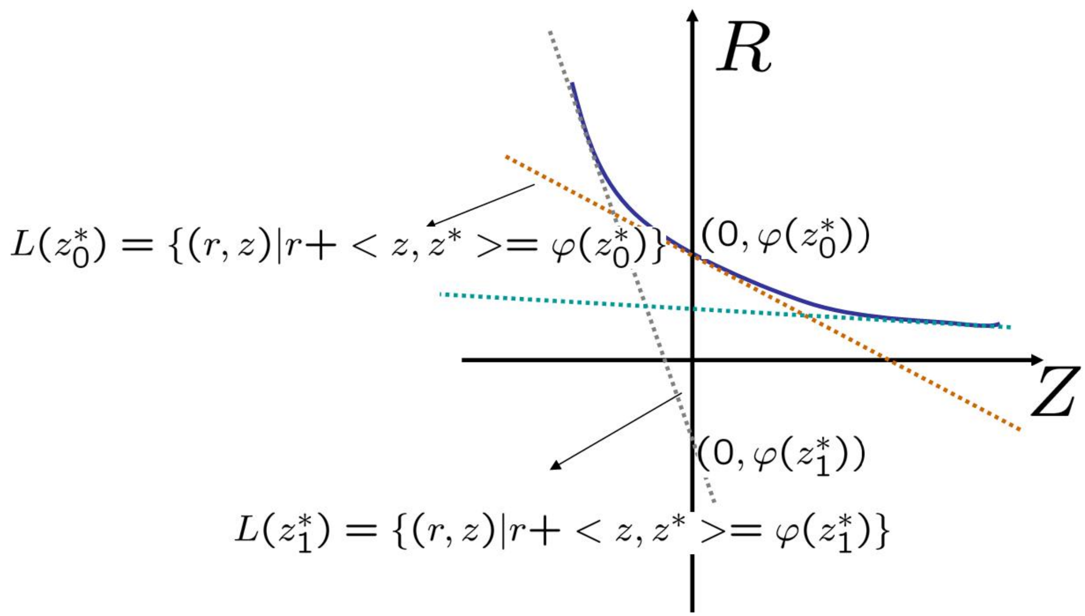

Figure 7: Shows supporting hyperplane to the epigraph of $\omega(z)$. Note that the tangents to the curve $\omega(z)$ are all negative from the fact that ω is a decreasing function. Also we are ruling out vertical hyperplanes. Thus each hyperplane can be described by $(1, z^*)$ with $z^* \geq 0$. Furthermore, each hyperplane has a y intercept of $\varphi(z^*)$ where $L(z^*) = \{(r,z)|r+ < z, z^* >= \varphi(z^*)\}$ describes the hyperplane. Also note that the maximum of these intercepts is $\omega(0)$. Thus one can postulate for convex problems that $\omega(0) = \max_{z^* \geq 0} \varphi(z^*)$.

The dual problem is defined on by first evaluating the dual function defined on the positive dual cone $P^{\oplus}$ given by

$$\varphi(z^*) := \inf\{L(x, z^*) : x \in \Omega\} = \inf_{x\in\Omega}\{f(x)+ < g(x), z^* >\}.$$

Theorem 40. *Let $z^* \in P^{\oplus}$ and $y^* \in Y$. Then*

$$\varphi(z^*) = \inf_{z\in\Gamma}\{\omega(z)+ < z, z^* >\}. \tag{42}$$

where Γ is the domain of the function $\omega : Z \to R$ that is
$\Gamma := \{z :$ *there exists $x \in \Omega$ such that $G(x) \leq z$*$\}$.

Proof: Let $z^* \geq 0$ and $z \in \Gamma$. Then

$$\begin{aligned}
\varphi(z^*) = \inf\{f(x) + < g(x), z^* >\} &\leq \inf\{f(x) + < g(x), z^* >: g(x) \leq z,\ x \in \Omega\} \\
&\leq \inf\{f(x) + < z, z^* >: g(x) \leq z,\ x \in \Omega\} \\
&= \inf\{f(x) : g(x) \leq z,\ x \in \Omega\} + < z, z^* > \\
&= \omega(z) + < z, z^* >
\end{aligned}$$

and thus

$$\varphi(z^*) \leq \inf_{z \in \Gamma}\{\omega(z) + < z, z^* > .$$

Suppose $x_1 \in \Omega$, let $z_1 = g(x_1)$. Then

$$\begin{aligned}
f(x_1) + < g(x_1), z^* > &\geq \inf\{f(x) + < z_1, z^* >: g(x) \leq z_1,\ z_1 = g(x_1),\ x \in \Omega\} \\
&\geq \inf\{f(x) + < z_1, z^* >: g(x) \leq z_1,\ \ x \in \Omega\} \\
&= \omega(z_1) + < z_1, z^* >
\end{aligned}$$

Thus

$$f(x_1)+ < g(x_1), z^* > \geq \inf_{z \in \Gamma} \{\omega(z)+ < z, z^* >\}.$$

Therefore

$$\varphi(z^*) = \inf\{f(x)+ < g(x), z^* >: x \in \Omega\} \geq \inf_{z \in \Gamma} \{\omega(z)+ < z, z^* >\}.$$

■

Consider the hyperplane defined by

$$< (r, z), (1, z^*) >= \varphi(z^*)$$

in $R \times Z$. Consider the set

$$A := \{(r, z) \in R \times \Gamma | r \geq \omega(z)\}$$

which is the epigraph $[\omega, \Gamma]$ of the function ω. Then clearly for all elements $(r, z_1) \in A$ we have

$$r+ < z_1, z^* >\geq \omega(z_1)+ < z_1, z^* >\geq \inf_{z\in\Gamma}\{\omega(z)+ < z, z^* >\} = \varphi(z^*).$$

Thus A is contained in the positive half space of the hyperplane

$$< (r, z), (1, z^*) >= \varphi(z^*).$$

It is also fairly clear that indeed $< (r, z), (1, z^*) >= \varphi(z^*)$ describes a supporting hyperplane as this hyperplane comes arbitrarily close to the epigraph of $\omega(z)$ given by $[\omega, \Gamma]$.

The above features of the hyperplane described by $< (r, z), (1, z^*) >= \varphi(z^*)$ is illustrated in Figure 7. Note that the $< (r, z), (1, z^*) >= \varphi(z^*)$ has a vertical intercept equal to $\varphi(z^*)$. Furthermore, it is evident from the Figure that $\omega(0)$ is the maximum of these intercepts. Thus one can postulate for convex problems that

$$\max_{z^* \geq 0} \varphi(z^*) = \omega(0) = \inf\{f(x) : x \in \Omega, g(x) \leq 0\}.$$

We will prove this next. It is interesting to note that Theorem 40 does not need convexity and thus the dual problem always provides a lower bound to the primal.

Theorem 41. *Consider the setup of Theorem 40 that has no requirements on convexity. Then*

$$\sup z^* \geq 0 \varphi(z^*) \leq \omega(0) = \inf\{f(x) : x \in \Omega, g(x) \leq 0\}.$$

Proof: From Theorem 40 we have that for any $z^* \geq 0$

$$\varphi(z^*) = \inf_{z \in \Gamma} \{\omega(z) + < z, z^* >\} \leq \omega(0).$$

Thus

$$\sup_{z^*\geq 0} \varphi(z^*) \leq \omega(0).$$

$\blacksquare$

Now we prove that

$$\max_{z^*\geq 0} \varphi(z^*) = \omega(0) = \inf\{f(x) : x \in \Omega, g(x) \leq 0\}$$

for the convex case.

Lagrange Duality Result

The KKT theorem states that for convex optimization problems the optimal primal value can be obtained via the following dual problem:

$$\max_{z^* \geq 0,\ y} \varphi(z^*, y).$$

The following is a Lagrange duality theorem.

Theorem 42. [Kuhn-Tucker-Lagrange duality] *Let X be a Banach space, Ω be a convex subset of X, Y be a finite dimensional normed space, Z be a normed space with positive cone P. Let Z^* denote the dual space of Z with a positive cone $P^{\oplus}$. Let $f : \Omega \to R$ be a real valued convex functional, $g : X \to Z$ be a convex mapping, $H : X \to Y$ be an affine linear map and $0 \in int[range(H)]$. Define*

$$\mu_0 := \inf\{f(x) : \ g(x) \leq 0,\ H(x) = 0,\ x \in \Omega\}. \tag{43}$$

Suppose there exists $x_1 \in \Omega$ *such that* $g(x_1) < 0$ *and* $H(x_1) = 0$ *and suppose* μ_0 *is finite. Then,*

$$\mu_0 = \max\{\varphi(z^*, y) : z^* \geq 0,\ z^* \in Z^*,\ y \in Y\}, \tag{44}$$

where $\varphi(z^*,\ y)$*:=* $\inf\{f(x)+ < g(x), z^* > + < H(x), y >: x \in \Omega\ \}$ *and the maximum is achieved for some* $z_0^* \geq 0$*,* $z_0^* \in Z^*,\ y_0 \in Y$*.*

Furthermore if the infimum in (43) is achieved by some $x_0 \in \Omega$ *then*

$$< g(x_0), z_0^* > + < H(x_0), y_0 >= 0, \tag{45}$$

and $x_0\ minimizes$

$$f(x)+ < g(x), z_0^* > + < H(x), y_0 >,\ \ over\ all\ x \in \Omega. \tag{46}$$

Proof: Given any $z^* \geq 0,\ y \in Y$ we have

$$\begin{aligned}\inf_{x\in\Omega}\{f(x)+ < g(x), z^* > \\ + < H(x), y >\} &\leq \inf_{x\in\Omega}\{f(x)+ < g(x), z^* > + < H(x), y > \\ &\quad :\ g(x) \leq 0,\ H(x) = 0\} \\ &\leq \inf_{x\in\Omega}\{f(x) :\ g(x) \leq 0,\ H(x) = 0\} \\ &= \mu_0.\end{aligned}$$

Therefore it follows that max$\{\varphi(z^*, y) : z^* \geq 0,\ y \in Y\} \leq \mu_0$. From Lemma 38 we know that there exists $z_0^* \in Z^*, z_0^* \geq 0,\ y_0 \in Y$ such that $\mu_0 = \varphi(z_0^*,\ y_0)$. This proves (44).

Suppose there exists $x_0 \in \Omega, H(x_0) = 0,\ g(x_0) \leq 0$ and $\mu_0 = f(x_0)$ then $\mu_0 = \varphi(z_0^*,\ y_0) \leq f(x_0)+ < g(x_0), z_0^* > + < H(x_0), y_0 > \leq f(x_0) = \mu_0$. Therefore we have $< g(x_0), z_0^* > + < H(x_0), y_0 >= 0$ and $\mu_0 = f(x_0)+ < g(x_0), z_0^* > + < H(x_0), y_0 >$. This proves the theorem. ∎

Sensitivity

Corollary 2. [Sensitivity] *Let X, Y, Z, f, H, g, Ω be as in Theorem 42. Let x_0 be the solution to the problem*

$$minimize\ f(x)$$

$$subject\ to\ x \in \Omega,\ H(x) = 0,\ g(x) \leq z_0$$

with (z_0^*, y_0) as the dual solution. Let x_1 be the solution to the problem

$$minimize\ f(x)$$

$$subject\ to\ x \in \Omega,\ H(x) = 0,\ g(x) \leq z_1$$

with (z_1^*, y_1) as the dual solution. Then,

$$< z_1 - z_0, z_1^* > \leq f(x_0) - f(x_1) \ \leq < z_1 - z_0, z_0^* > . \tag{47}$$

Proof: From Theorem 42 we know that for any $x \in \Omega$,

$$\begin{aligned} & f(x_0)+ < g(x_0) - z_0, z_0^* > + < H(x_0), y_0 > \\ \leq \quad & f(x)+ < g(x) - z_0, z_0^* > + < H(x), y_0 > . \end{aligned}$$

In particular we have

$$\begin{aligned} & f(x_0)+ < g(x_0) - z_0, z_0^* > + < H(x_0), y_0 > \\ \leq \quad & f(x_1)+ < g(x_1) - z_0, z_0^* > + < H(x_1), y_0 > . \end{aligned}$$

From Theorem 42 we know that $< g(x_0) - z_0, z_0^* > + < H(x_0), y_0 >= 0$ and $H(x_1) = 0$. This implies

$$f(x_0) - f(x_1) \ \leq < g(x_1) - z_0, z_0^* > \leq < z_1 - z_0, z_0^* > .$$

A similar argument gives the other inequality. This proves the corollary. ■