1 Introduction

What is and Why Non smooth Optimization=Essential Tool and Analysis

- Nonsmooth variational problem (Material Science, Finance and Free boundary, Nonlinear Equilibrium)

$$\min \quad \int_0^1 W(u_x(x)) + j(u(x)) - u(x)f(x)\,dx \text{ subject to } u(x) \leq \psi(x), \quad x \in \Omega = (0,1).$$

  over $u \in X = H_0^1(0,1)$, a proper function space, where W : energy functional, j : potential functional and ψ : pointwise obstacle function. For example

$$W(u_x) = \frac{1}{p}|u_x|^p + \frac{\alpha}{2}|u_x|^2,\ 0 \leq p \leq \infty \text{ and } j(u) = \frac{1}{4}(u^2-1)^2,\ |u| \ \text{ (phase field)}$$

  But, also $W : R \to R^+$ and $j : R \to R$ are non convex and non differentiable. Objective: Existence, Uniqueness and Regularity, Find a contact region $\{x \in \Omega : u(x) = \psi(x)\}$. Effective Numerical method and Algorithm.

- Nonsmooth regression: Least 1-norm (LASSO):

$$\min |Au - b|_1 + \alpha\, |u|_1 \text{ over } u \in R^n$$

  Or, $|x|_0$=number of nonzero entry (0-metric). A is a convolution, scattering matrix and can be ill-posed $\to$ ill-posed inverse problem.

- Nonsmooth classification: Sparse support vector machine:

$$\min \quad |(g(x_1)-1)^+|_1 + |(g(x_2)+1)^-|_1 + \beta\, |w|_1 \text{ over } (w,b).$$

  where $x_1 \in \Omega^+$ and $x_2 \in \Omega^-$ and $g(x) = w \cdot x + b, \quad w \in R^d, \quad b \in R$ is an affine decision function. — Kernel machine, AI classification (Large Scale).

- Optimal control and Inverse problems:

$$\min J(y,u) = F(y) + H(u) \text{ subject to } E(y,u) = 0 (\text{ Equality constraint (PDEs constrain)})$$

  over $y \in \mathcal{C} \subset X$ (state constraint) and $u \in \mathcal{K} \subset U$ (control constraint). — Uncertainty, Risk, Eextreme (Rare) event.

Lectures:

- Function space and Basic Functional analysis
- Hilbert space theory, Constrained optimization and Constrained Qualification.
- Convex analysis, Variational inequalities and Dual problem.
- Lagrange multiplier theory for nonsmooth optimization and Sensitivity analysis

- Augmented Lagrangian method and Primal Dual Active set method.
- Penalty method and Semi-smooth Newton method (Algorithms, Convergence Analysis)
- Examples and Applications (Constrained optimization, Bayes and (AI) machine learning, Image analysis).
- Advanced Topics: HJB (Hamilton-Jacobi-Bellman), Mean-Field game, MPEC and Mass Transport, Finance.

Reference: Ito-Kunish (Optimization) and Ito-Jin book (Inverse problem) and Online-lecture note

Basic Function Space Theory

- Vector Space, Normed space, Banach space, Hilbert space.
- Dual space and Hahn-Banach theorem, Weak and Weak star Convergence, Riesz representation theorem.
- Linear Operator Theory, Open mapping, Uniform bounded principle, closed range theory, compact operator.
- Distribution and Generalized Derivatives, Sobolev space.
- Lax-Milgram Theorem ann Banach-Necas-Babuska Theorem.

Optimization Theory

- Hilbert (Banach) space Theory, Riesz Representation, Minimum norm problem
- Variational inequalities and Non-smooth optimization
- Lagrange Multiplier Theory for Constrained optimizations, Saddle point problems
- Penalty method and Generalized Saddle point problems
- Lagrange calculus and Sensitivity Analysis
- Convex analysis and Nonsmooth convex optimization
- Augmented Lagrangian method and Dual optimization
- Semi-smooth Newton method and Primal dual Active set method

Notation. We consider Real number field R. Dual product on $X \times X$ is defined by

$$x^*(x) = \langle x, x^* \rangle_{X \times X^*} \text{ or } \langle x^*, x \rangle_{X^* \times X}.$$

for $x \in X$, a norm space $(X, |\cdot|)$ and $x^* \in X^*$, a bounded linear functional on X

Let us denote by $F : X \to X^*$, the duality mapping of X, i.e.,

$$F(x) = \{x^* \in X^* : \langle x, x^* \rangle = |x|^2 = |x^*|^2\}.$$

By Hahn-Banach theorem, $F(x)$ is non-empty. In general F is multi-valued. Therefore, when X is a Hilbert space, $\langle \cdot, \cdot \rangle$ coincides with its inner product if X^* is identified with X and $F(x) = x$.

For $X = W_0^{1,p}(\Omega)$ with norm

$$|u|_X = (\int_\Omega |\nabla u|^p \, dx)^{1/p}$$

Then, $X^* = W^{-1,q}(\Omega)$, $\frac{1}{p} + \frac{1}{q} = 1$ and

$$F(u) \sim -\nabla \cdot (|\nabla u|^{p-2} \nabla u) \in X^*$$

since

$$\langle F(u), u \rangle = \int_\Omega |\nabla u|^p \, dx$$

For $f \in X^*$, $u \sim F^{-1}(f) \in X$ is given by

$$-\nabla \cdot (|\nabla u|^{p-2} \nabla u) = f \text{ in } X^*$$

i.e.

$$\int_\Omega (\nabla u|^{p-2} \nabla u, \nabla \phi) \, dx = \langle f, \phi \rangle \text{ for all } \phi \in X$$

and it minimizes

$$\frac{1}{p} \int_\Omega |\nabla u|^p \, dx - \langle f, u \rangle \text{ over } X = W_0^{1,p}(\Omega).$$

Also, we have $F = \partial \frac{1}{2} |\cdot|_X^2$ (sub-differtial of norm):

$$F(x) = \{f \in X^* : \frac{1}{2}|y|^2 - \frac{1}{2}|x|^2 \geq \langle f, y - x \rangle \text{ for all } y\}.$$

Existence. min $J(x)$ over $x \in \mathcal{C}$. J: Coercive, weakly (weakly star) compact, and weakly (weakly star) lower sequentially semi-continuous $\to$ Existence. (Weirestrass Theorem).

$$J(x) \to +\infty \text{ as } |x|_X \to +\infty.$$

Let x_n be a minimizing sequence such that $J(x_n)$ is decreasing and $\lim_{n\to\infty} J(x_n) = \eta = \inf_{x\in\mathcal{C}} J(x)$.

$$\liminf_{n\to\infty} J(x_n) \le J(x) \text{ for all } x_n \text{ that weakly converges to } x \text{ in } X.$$

Orthogonal decomposition $N(E) = R(E^*)^\perp$ and

$$X = N(E) \oplus \overline{R(E^*)}.$$

Closed Range Theorem.

Mathematical programing.

$$\min \quad J(x) \quad \text{subject to } E(x) = 0 \text{ and } G(x) \le 0. \tag{1.1}$$

over a closed convex set $\mathcal{C}$ in X.

Lagrange Multiplier Theory

$$L(x,\lambda) = F(x) + \langle \lambda, E(x) \rangle$$

Consider the max-min problem

$$\max_\lambda \min_x L(x,\lambda)$$

Define $V(\lambda) = \min_x \; L(x,\lambda)$. Then, $V'(\lambda) = E(x)$. We obtain

$$\begin{cases} F'(x) + E'(x)^*\lambda = 0 \\ \\ E(x) = 0 \end{cases}$$

Augmented Lagrangian: For $c > 0$

$$L_c(x, \lambda) = F(x) + \langle \lambda, E(x) \rangle + \frac{c}{2}|E(x)|^2$$

$$\begin{cases} F'(x_c) + E'(x_c)^*(\lambda + \lambda_c) = 0 \\ \\ \frac{1}{c}\,\lambda_c = E(x_c). \end{cases}$$

Convex Non-smooth optimization

$$\min_{x \in \mathcal{C}} \; F(x) + \varphi(\Lambda x) \tag{1.2}$$

where F is C^1, φ is convex on H and $\Lambda \in \mathcal{L}(X, H)$.

$$F(x) + \varphi(\Lambda x - u) + (\lambda, u) \text{ for } u = 0$$

$$\begin{cases} F'(x) + \Lambda^* \lambda = 0 \\ \\ \lambda \in \partial\varphi(\Lambda x) \end{cases}$$

$$\min_{x \in C, v \in H} \quad f(x) + \varphi(\Lambda x - v) + (\lambda, v)_H + \frac{c}{2}\,|v|_H^2, \tag{1.3}$$

Minimizing over u gives

$$\min_{y \in C} \quad L_c(y, \lambda) = f(x) + \varphi_c(\Lambda x, \lambda). \tag{1.4}$$

where for $u = \Lambda x$

$$\varphi_c(u, \lambda) = \inf_{v \in H} \{\varphi(u - v) + (\lambda,\, v)_H + \tfrac{c}{2}\,|v|_H^2\}$$

$$= \inf_{z \in H} \{\varphi(z) + (\lambda, u - z)_H + \frac{c}{2}\,|u - z|_H^2\} \;\; (u - v = z). \tag{1.5}$$

where $\varphi_c(u, \lambda)$ is called the generalized Yoshida-Moreau approximations of φ. Then, the augmented Lagrangian functional of (1.2) is given by

$$L_c(x, \lambda) = f(x) + \varphi_c(\Lambda x, \lambda). \tag{1.6}$$

Let φ^* is the convex conjugate of φ

$$\varphi_c(u,\lambda) = \inf_{z\in H}\{\sup_{y\in H}\{(y,z)-\varphi^*(y)\}+(\lambda,u-z)_H+\tfrac{c}{2}\,|u-z|_H^2\}$$

$$= \sup_{y\in H}\{-\frac{1}{2c}|y-\lambda|^2+(y,u)-\varphi^*(y)\}. \tag{1.7}$$

where we used

$$\inf_{z\in H}\{(y,z)_H+(\lambda,u-z)_H+\frac{c}{2}\,|u-z|_H^2\} = (y,u)-\frac{1}{2c}|y-\lambda|_H^2$$

Thus, from Theorem 2.2

$$\varphi_c'(u,\lambda) = y_c(u,\lambda) = \mathrm{argmax}_{y\in H}\{-\tfrac{1}{2c}|y-\lambda|^2+(y,u)-\varphi^*(y)\}$$

$$= \lambda + c\,v_c(u,\lambda) \tag{1.8}$$

where

$$v_c(u,\lambda) = \mathrm{argmin}_{v\in H}\{\varphi^*(u-v)+(\lambda,v)+\frac{c}{2}|v|^2\} = \frac{\partial}{\partial\lambda}\varphi_c(u,\lambda)$$

Moreover, if

$$p_c(\lambda) = \mathrm{argmin}_{p\in H}\{\frac{1}{2c}|p-\lambda|^2+\varphi^*(p)\} = \mathrm{prox}_{\varphi^*}(\lambda),, \tag{1.9}$$

then we have

$$\varphi_c'(u,\lambda) = p_c(\lambda+c\,u). \tag{1.10}$$

Theorem (Lipschitz complementarity)

(1) If $\lambda\in\partial\varphi(x)$ for $x,\ \lambda\in H$, then $\lambda=\varphi_c'(x,\lambda)$ for all $c>0$.
(2) Conversely, if $\lambda=\varphi_c'(x,\lambda)$ for some $c>0$, then $\lambda\in\partial\varphi(x)$.

Proof: If $\lambda\in\partial\varphi(x)$, then from (1.5) and (1.7)

$$\varphi(x)\ge\varphi_c(x,\lambda)\ge\langle\lambda,x\rangle-\varphi^*(\lambda)=\varphi(x).$$

Thus, $\lambda \in H$ attains the supremum of (1.8) and we have $\lambda = \varphi_c'(x,\lambda)$. Conversely, if $\lambda \in H$ satisfies $\lambda = \varphi_c'(x,\lambda)$ for some $c > 0$, then $v_c(x,\lambda) = 0$ by (1.8). Hence it follows from that

$$\varphi(x) = \varphi_c(x,\lambda) = (\lambda, x) - \varphi^*(\lambda).$$

which implies $\lambda \in \partial\varphi(x)$. □

$$\begin{cases} \langle f'(\bar{x}) + \Lambda^*\bar{\lambda}, x - \bar{x}\rangle_{X^*,X} \geq 0, \quad \text{for all } x \in \mathcal{C} \\ \\ \lambda = \varphi_c'(\Lambda\bar{x}, \bar{\lambda}). \end{cases} \tag{1.11}$$

which is for the primal-dual variable $(\bar{x}, \bar{\lambda})$. The advantage here is that the frequently employed differential inclusion $\bar{\lambda} \in \partial\varphi(\Lambda\bar{x})$ is replaced by the equivalent nonlinear equation (1.11).

<u>Time-dependent problem</u>

$$\min \quad \int_0^T (\ell(x(t)) + h(u(t))\, dt + G(x(T))$$

subject to

$$\frac{d}{dt}x(t) = Ax(t) + F(x(t)) + Bu(t), \;\; x(0) = x, \;\; u(t) \in U$$

(Implicit) Time-Discretization:

$$\frac{x^n - x^{n-1}}{\Delta t} - Ax^n - F(x^n) = Bu^n$$

Well-posedness: Find $x^n \in X$ given $x^{n-1} + \Delta t\, Bu^n \in X^*$ for

$$x^n - \Delta t\, \tilde{A}(x^n) = x^{n-1} + \Delta t\, Bu^n$$

Lax-Milgram, Minty-Browder, Pseudo-Monotone operator theory.

For $(\vec{x},\vec{u},\vec{p}) \in X^N \times U^N \times X^N$ define

$L(\vec{x},\vec{u},\vec{p}) =$

$\sum_n (-\frac{x^n - x^{n-1}}{\Delta t} + Ax^n + F(x^n) + Bu^n, p^n) + \ell(x^n) + h(u^n))\Delta t + G(x^N)$

Multiplier Theory says

$$\begin{cases} -\frac{p^n - p^{n+1}}{\Delta t} = (A + F'(x^n))^* p^n + \ell'(x^n) = 0, \quad p^N = G'(x^N) \\ \\ -B^* p^n \in \partial h(u^n) \end{cases}$$

CONVEX ANALYSIS AND VARIATIONAL PROBLEMS by IVAR EKELAND and ROGER TEMAM

p-Laplacian problem

$$\Phi(x,y) = f(x) + \varphi(\Lambda x + y) \tag{1.12}$$

$$(P) \quad \inf_x \Phi(x,0) \quad (P^*) \quad \sup_{y^*} -\Phi^*(0,y^*)$$

$$\Phi^*(x^*,y^*) = f^*(x^* - \Lambda^* y^*) + \varphi^*(y^*).$$

Primal and Dual problem

$$(P) \quad \min \quad J(u) = \int_\Omega \{W(|\nabla u|^2) - (f,u)\}\, dx$$

(P) is equivalent to

$$-\nabla \cdot (W'(|\nabla u|^2)\nabla u) - f = 0$$

Monotone convergent fixed point method:

$$-\nabla \cdot (W'(|\nabla u^n|^2)\nabla u^{n+1}) - f = 0$$

if $s \to W(s)$ is concave ($p \leq 2$). Multiplying this by $u^{n+1} - u^n$,

$$\int_\Omega W'(|\nabla u^n|^2)(\nabla u^{n+1}), \nabla u^{n+1} - \nabla u^n) - (f, u^{n+1} - u^n) = 0$$

where

$$(\nabla u^{n+1}, \nabla u^{n+1} - \nabla u^n) = \tfrac{1}{2}(|\nabla u^{n+1}|^2 - \nabla u^n)|^2 + |\nabla u^{n+1} - |\nabla u^n)|^2)$$

$$W(|\nabla u^{n+1}|^2) - W(|\nabla u^n|^2) - W'(|\nabla u^n|^2)(|\nabla u^{n+1} - |\nabla u^n)|^2) \leq 0$$

Thus, we obtain

$$J(u^{n+1}) + W'(|\nabla u^n|^2)|\nabla u^{n+1} - \nabla u^n|^2 \leq J(u^n).$$

If $W(t^2) = \frac{1}{p}\,|t|^p$, then

$$(P^*) \quad \sup\ -\frac{1}{q}\int_\Omega |\vec{p}|^q\,dx \text{ subject to } \nabla\cdot\vec{p} = f$$

Let $\vec{p} = \nabla\phi + \operatorname{curl}\psi$ with $\Delta\phi = f$. In fact, let

$$F(u) = -\int_\Omega f u\,dx, \quad \varphi(v) = \int_\Omega w(v^2)\,dx, \quad \Lambda = \nabla$$

Then,

$$F^*(-\Lambda^*\vec{p}) = I_{\nabla\vec{p}=f} \text{ and } \varphi^*(\vec{p}) = \int_\Omega w^*(\vec{p})\,dx$$

Define $\vec{p} = \nabla\phi + \operatorname{curl}\psi$. Then (P^*) is equivalent to

$$\min \quad \int_\Omega |\nabla\psi + \operatorname{curl}\phi|^q\,dx.$$

$$\vec{p} = -W'(|\nabla u|^2)\nabla u$$

<u>Ł^0 minimization</u> Consider the minimization on H;

$$\min \quad J(y) + N(y), \quad N(y) = \int_\Omega h(y(\omega))\,d\omega. \qquad (1.13)$$

$$\begin{cases} J'(u)+\lambda=0 \\ \\ u\in\Phi(\lambda) \end{cases}$$

where

$$\Phi(q) := argmin_{u\in R}(h(u)-qu) = \begin{cases} \frac{q}{\alpha} & \text{for } |q|\geq\sqrt{2\alpha\beta} \\ \\ 0 & \text{for } |q|<\sqrt{2\alpha\beta.} \end{cases} \tag{1.14}$$

and the conjugate function h^* of $h=\frac{\alpha}{2}|u|^2+\beta\,|u|_0$;

$$-h^*(q)=h(\Phi(q))-q\,\Phi(q)=\begin{cases} -\frac{1}{2\alpha}|q|^2+\beta & \text{for } |q|\geq\sqrt{2\alpha\beta} \\ \\ 0 & \text{for } |q|<\sqrt{2\alpha\beta.} \end{cases}$$

∂h^* is the maximal monte extension of Φ. The bi-conjugate function $h^{**}(u)$ is the convexfication of h and is given by

$$h^{**}(u)=\begin{cases} \frac{\alpha}{2}|u|^2+\beta \quad |u|>\sqrt{\frac{2\beta}{\alpha}} \\ \\ \sqrt{2\alpha\beta}\,|u| \quad |u|\leq\sqrt{\frac{2\beta}{\alpha}}. \end{cases}$$

$$u\in\partial h^*(\lambda)\Leftrightarrow\lambda\in\partial h^{**}(u)$$

Let $\mu=\alpha\,u+p$ and $p=J'(u)$. Then complementarity is given by

$$\begin{cases} \mu=0 & \text{if } |\mu-\alpha\,u|>\sqrt{2\alpha\beta} \\ \\ u=0 & \text{if } |\mu-\alpha\,u|\leq\sqrt{2\alpha\beta} \end{cases}$$

1. Initialize $u^0\in X$ and $\lambda^0\in H$. Set $n=0$.
2. Define the active index and inactive index by

$$\mathcal{A}=\{k:|\mu_k-\alpha\,u_k|\leq\sqrt{2\alpha\beta}\},\quad \mathcal{I}=\{j:|\mu_j-\alpha\,u_j|>\sqrt{2\alpha\beta}\}$$

3. Let $\lambda^+ = \frac{u}{\alpha}$ on $\mathcal{I}$ and $u^+ = 0$ on $\mathcal{A}$.

4. Solve for (u^+, λ^+)

$$J'(u) + \lambda^+ = 0$$

5. Stop or set $n = n + 1$ and return to step 1,

ℓ^p minimization Consider

$$\min \quad F(x) + \frac{\beta}{p}\,|x|^p$$

Regularized $|t|^p \sim \Psi_\epsilon(t^2)$ for $\epsilon > 0$

$$\Psi_\epsilon(t) = \begin{cases} \frac{p}{2}\frac{t}{\epsilon^{2-p}} + (1 - \frac{p}{2})\epsilon^p & \text{for } 0 \le t \le \epsilon^2 \\ \\ t^{\frac{p}{2}} & \text{for } t \ge \epsilon^2 \end{cases}$$

It results in the fixed point iterate of the form

$$F'(x^{n+1}) + \frac{\beta}{\max(\epsilon^{2-p}, |x^n|^{2-p})} x^{n+1} = 0$$

OPTIMIZATION BY VECTOR SPACE METHODS by David G. Luenberger

Definition. The vectors $x \in X$ and $x^* \in X^*$ are said to be orthogonal if $\langle x, x^* \rangle = 0$. $M^\perp = \{x^* \in X^* : \langle m, x^* \rangle = 0\}$ for all $m \in M\}$.

$$[M^\perp]^\perp = M \quad (M\text{: closed subspace of } X)$$

Theorem 1. Let x be an element in a real normed linear space X and let d denote its distance from the subspace M. Then,

$$d = \inf_{m \in M} |x - m| = \max_{x^* \in M^\perp,\ |x^*| \le 1} \langle x, x^* \rangle, \tag{1.15}$$

where the maximum on the right is achieved for some x_0^*. If the infimum on the left is achieved for some $m_0 \in M$ then x_0^* is aligned with $x - m_0$.

Proof: For $\epsilon > 0$, let $m_\epsilon \in M$ satisfy $|x - m_\epsilon| \leq d + \epsilon$. Then for any $x^* \in M^\perp$and $|x^*| \leq 1$, we have

$$\langle x, x^* \rangle = \langle x - m_\epsilon, x^* \rangle \leq |x^*||x - m_\epsilon|$$

Since $\epsilon > 0$ is arbitrary, we conclude that $\langle x, x^* \rangle \leq d$ Thus, the proof of the first part of the theorem is complete if we find any x_0^* such that $\langle x, x_0^* \rangle = d$. Let S be the subspace $[x + M]$, i.e., elements of S are uniquely represented in the form $s = \alpha\, x + m$, with $m \in M$ and $\alpha \in R$. Define the linear functional f on S by $f(s)\alpha\, d$. We have

$$|f| = \sup_{s \in S} \frac{|f(s)|}{|s|} = \sup \frac{|\alpha|\, d}{|\alpha x + m|} = \sup \frac{|\alpha|\, d}{|\alpha||x + \frac{m}{\alpha}|} = \frac{d}{\inf |x + \frac{m}{\alpha}|} = 1$$

It follows from the Hahn-Banach extension, there exists an extension $x_0^* \in X^*$ of f from S to X such that $|x_0^*| = 1$ and $x_0^* = f$ on S. By construction, we have $x_0^* \in M^\perp$ and $\langle s, x_0^* \rangle = d$. which complete the first part of the theorem.

Now, assume that there is an $m_0 \in M$ with $|x - m_0| = d$. For x_0^* constructed above Then

$$\langle x - m_0, x_0^* \rangle = d = |x_0^*||x - m_0|.$$

and x_0^* is aligned with $x - m_0$. $\square$

Geometrically (1) implies that the error $x - m_0$ is orthogonal to M. Theorem states the equivalence of two optimization problems: one in X called the primal problem and the other in X^* called the dual problem. The problems are related through both the optimal values of their respective objective functionals and an alignment condition on their solution vectors. Since in many spaces alignment can be explicitly characterized, the solution of either problem often leads directly to the solution of the other. Duality relations such as this are therefore often of extreme practical as well as theoretical significance in optimization problems.

Corollary 1 (generalization of the projection theorem) Let x be an element of a real normed linear vector space X and let M be a subspace of X. A vector $m_0 \in M$ satisfies $|x - m_0| \leq |x - m|$ for all $m \in M$ if and only if there is a nonzero vector $x_0^* \in M^\perp$ aligned with $x - m_0$.

Proof: To prove the "if" part, assume that $x - m_0$ is is aligned with an $x^* \in M^\perp$. Then we have

$$\langle x, x^* \rangle = \langle x - m, x^* \rangle \leq |x - m| \text{ for all } m \in M$$

where

$$\langle x, x^* \rangle = \langle x - m_0, x^* \rangle = |x - m|.$$

As a companion to Theorem 1, we have:

Theorem 2 (dual). Let M be a subspace in a real normed space X. For $x^* \in X^*$

$$d = \min_{m^* \in M^\perp} |x^* - m^*| = \sup_{|x| \leq 1,\, x \in M} \langle x, x^* \rangle. \tag{1.16}$$

where the minimum on the left is achieved for $m_0 \in M^\perp$. If the supremum on the right is achieved for some $x_0 \in M$ then $x^* - m_0^*$ is aligned with x_0.

Minimum norm solution. Let $M = span\{y_i\}$. If $\bar{x}^*$ is any vector satisfying the constraints, we have

$$d = \min_{\langle y_i, x^* \rangle = c_i} |x^*| = \min_{m^* \in M^\perp} |\bar{x}^* - m^*|.$$

Since

$$d = \min_{m^* \in M^\perp} |\bar{x}^* - m^*| = \sup_{|x| \leq 1,\, x \in M} \langle x, \bar{x}^* \rangle,$$

we have

$$d = \min_{\langle y_i, x^* \rangle = c_i} |x^*| = \max_{|Ya| \le 1} \langle Ya, \bar{x}^* \rangle = \max_{|Ya| \le 1} c^t a \qquad (1.17)$$

since any vector in M is of the form $m = \sum_i a_i y_i = Ya$ where $Y \in \mathcal{L}(R^n, X)$ and $Y^* x^* = c$.

Example 2 (A Control Problem) Consider the problem of selecting the field current $u(t)$ on $[0,1]$ to drive a motor governed by

$$\theta''(t) + \theta'(t) = u(t)$$

from the initial conditions $\theta(0) = \theta'(0) = 0$ to $\theta(1) = 1,\ \theta'(1) = 0$ such a way as to minimize

$$\max_{t \in [0,1]} |u(t)|.$$

This example is similar to minimizing the total energy

$$\min \quad \int_0^1 |u(t)|^2 \, dt$$

but now our objective function reflects a concern with possible damage due to excessive current. The problem can be thought of as being formulated in $C[0,1]$, but since this is not the dual of any normed space, we are not guaranteed that a solution exists in $C[0,1]$. Thus, instead we take $X = L^1(0,1),\ X^* = L^\infty(0,1)$ and seek $u \in X^*$ of minimum norm. The terminal constraints are given as

$$\textstyle\int_0^1 e^{t-1} u(t) \, dt = 0$$

$$\textstyle\int_0^1 (1 - e^{t-1}) u(t) \, dt = 1.$$

Thus, from (1.17) with $X = L^1(0,1)$

$$\max \quad a_2 \text{ subject to } \int_0^1 |(a_1 - a_2) e^{t-1} + a_2| \, dt \le 1$$

The general nature of the optimal control is easily deduced from the alignment requirement. Obviously, the function $(a_1 - a_2)e^{t-1} + a_2$ being the sum of a constant and an exponential term, can change sign at most once, and since the optimal $u \in X^*$ is aligned with this function, u must be "bang-bang" (Le., it must have values $\pm M$ for some M) and changes sign at most once, i.e.

$$u = M\text{sign}((a_1 - a_2)e^{t-1} + a_2).$$

Example 3. (Rocket Problem) Consider the problem of selecting the thrust program $u(t)$ for a vertically ascending rocket-propelled vehicle, subject only to the forces of gravity and rocket thrust in order to reach a given altitude with minimum fuel expenditure. Assuming fixed unit mass, unit gravity, and zero initial conditions, the altitude x(t) is governed by a differential equation of the form

$$x''(t) = u(t) - 1$$

with initial conditions $x(0) = x'(0) = 0$. Then,

$$x(T) = \int_0^T (T - t)u(t)\,dt - \frac{T^2}{2}$$

Our problem is to attain a given altitude, say $x(T) = 1$, while minimizing the fuel expense.

$$\int_0^T |u(t)|\,dt$$

The final time T is in general unspecified, but we approach the problem by finding the minimum fuel expenditure for each fixed T and then minimizing over T. For a fixed T the optimization problem reduces to that of finding u minimizing $L^1(0,T)$ norm of u subject to $x(T) = 1$. Since, however, $L^1(0,T)$ is not the dual of any normed space, we imbed our problem in the space $BV[0,T]$ and associate control elements u with the derivatives of elements v in $BV[0,T]$. Thus the problem becomes that of finding $v \in NBV[0,T]$ minimizing

$$\int_0^T |dv(t)| = |v|_{X^*},$$

where $X = C[0,T]$ and $X^* = BV[0,T]$, subject to

$$\int_0^T (T-t)dv(t) = 1 + \frac{T^2}{2}.$$

From (1.17)

$$\min \ \ |v|_{BV[0,T]} = \max_{|a(T-t)|_{C[0,T]} \leq 1} a(1 + \frac{T^2}{2}).$$

Since

$$|(T-t)a|_{C[0,T]} = max_{t\in[0,T]}|(T-t)a| = T|a|.$$

where the maximum is attained at $t = 0$. The optimal v must be aligned with $(T-t)a$ and, hence, can vary only at $t = 0$. Therefore, we conclude that v is a step function and u is an impulse (or delta function) at $t = 0$. The best final time can be obtained by differentiating the optimal fuel expenditure with respect to T. This leads to the final result $T = \sqrt{2}$ and $\min |v|_{BV} = \sqrt{2}$. Note that our early observation that the problem should be formulated in $BV[0,T]$ rather than $L^1(0,T)$ turned out to be crucial since the optimal u is an impulse.

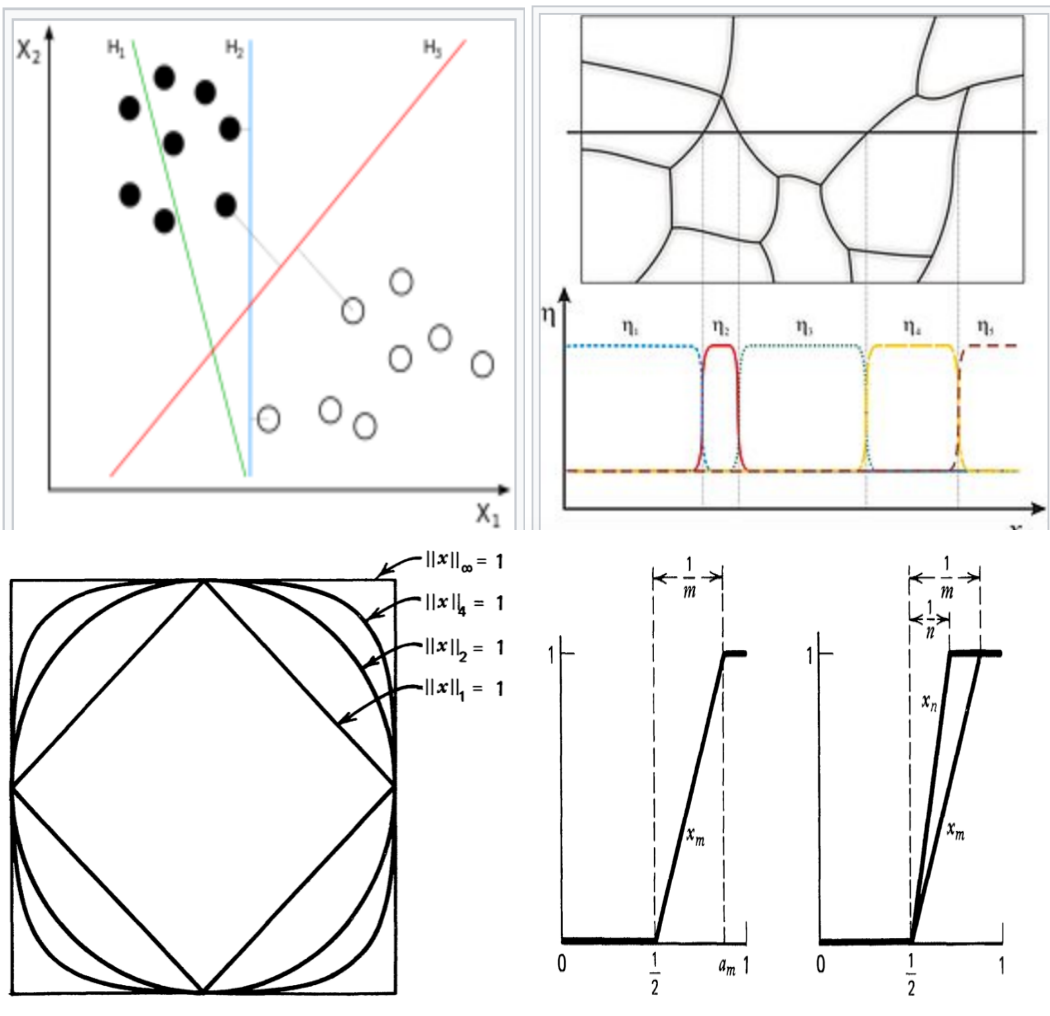

Fig. 16. Unit spheres in Prob. 10

Figure 1: Case: Vector Support Machine, Phase Field Model, Unit ball, Nonconverget Cauchy sequence

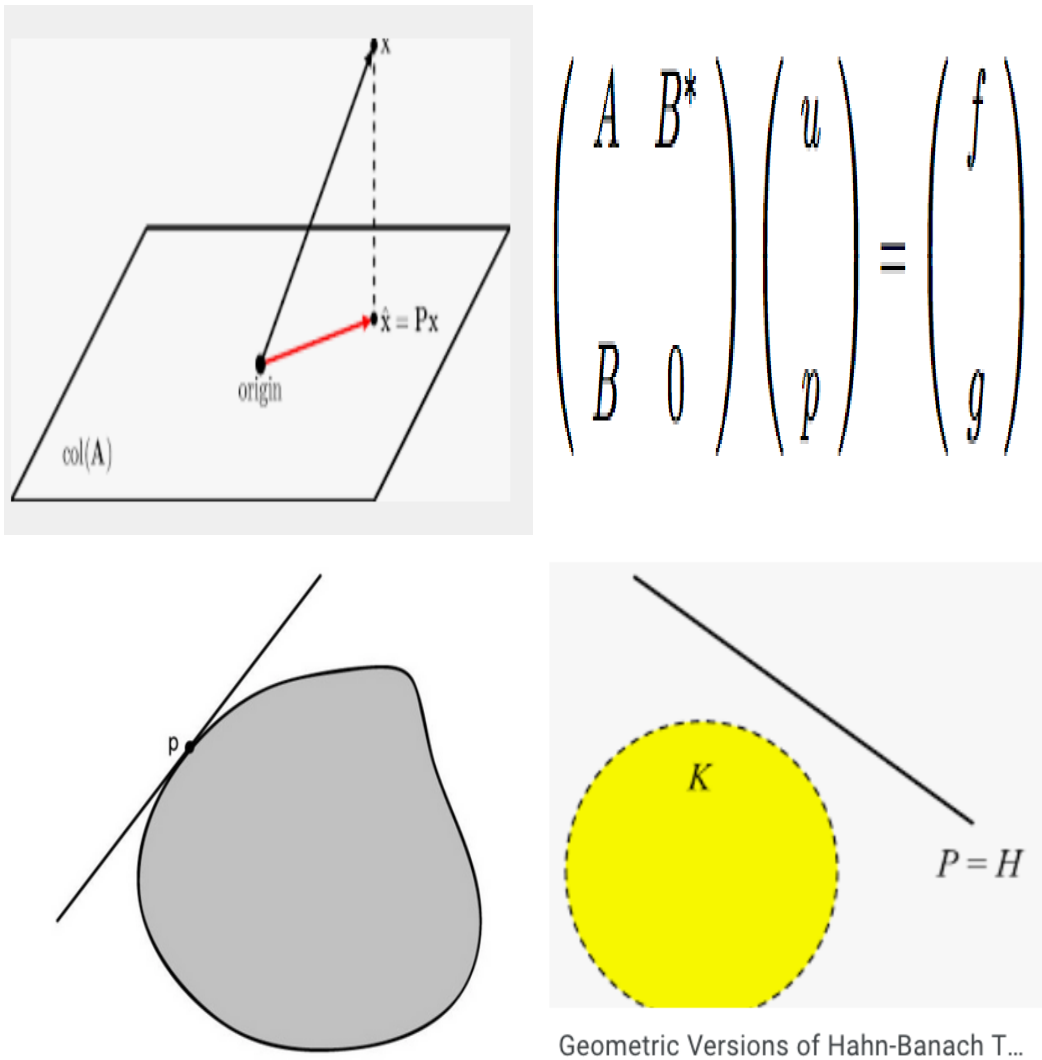

Figure 2: Orthogonal Projection, Saddle problem