Chapter 1
Mathematical basics and convex analysis

Convex Optimization for Computer Vision
SS 2017

Michael Moeller
Visual Scene Analysis
Department of Computer Science
and Electrical Engineering
University of Siegen

Mathematical basics

Notation, norm, inner product

We will work in the vector space $\mathbb{R}^n$ equipped with an inner product

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

for $x, y \in \mathbb{R}^n$.

The ℓ^2 norm is *induced* by this inner product, i.e.

$$\|x\|_2 = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

If I omit the subscript and just write $\|x\|$, I mean the ℓ^2 norm. If other norms are meant, I will make it explicit by subscripts, e.g. the ℓ^1 or the ℓ^∞ norms

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad , \qquad \|x\|_\infty = \max_i |x_i|$$

Topological properties of sets

Definitions and things to recall from analysis 1:

Definitions

- The (open) ℓ^2-ball of radius ϵ around some $x \in \mathbb{R}^n$ is defined as

$$B(x, \epsilon) = \{y \in \mathbb{R}^n \mid \|y - x\|_2 < \epsilon\}$$

- A set $C \subset \mathbb{R}^n$ is called **open** if for all $x \in C$ there exists a $\epsilon > 0$ such that the ball with radius ϵ around x, $B(x, \epsilon)$, is contained in C: $B(x, \epsilon) \subset C$.
- A set $C \subset \mathbb{R}^n$ is called **closed** if its complement $C^c = \{x \in \mathbb{R}^n \mid x \notin C\}$ is open.

A set is closed if and only if it contains all its limit points!

Board: Some examples

Topological properties of sets

Important further concepts to talk about sets:

Definitions

- The **closure** of a set $C \subset \mathbb{R}^n$ is

$$\overline{C} = \{x \mid \text{ there exists a convergent sequence } (x_n)_n \subset C \text{ such that } \lim_{n \to \infty} x_n = x\}$$

- The **interior** of a set $C \subset \mathbb{R}^n$ is

$$\mathring{C} = \{x \in C \mid \text{ there exists } \epsilon > 0 \text{ such that } B(x, \epsilon) \subset C\}$$

Further examples on the board!

Further properties of sets

We will need the concept of boundedness and compactness

Definitions

- A set $C \subset \mathbb{R}^n$ is called bounded if there exists an $r > 0$ such that $C \subset B(0, r)$, i.e. $\|x\|_2 \leq r \;\; \forall x \in C$.
- A set $C \subset \mathbb{R}^n$ is called compact if it is closed and bounded.

Bolzano-Weierstrass Theorem

Let $(x_n)_{n\in\mathbb{N}} \subset C$ be a sequence in the compact set C. Then there exists a convergent subsequence (x_{n_k}) and $\hat{x} := \lim_{k\to\infty} x_{n_k} \in C$.

We will use this theorem for showing the existence of minimizers to a class of convex optimization problems.

Continuity of functions

Definition: Convergent sequences

We say that a sequence $(x_n) \subset \mathbb{R}^n$ converges to $\hat{x} \in \mathbb{R}^n$ if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$x_n \in B(a, \epsilon) \qquad \forall n \geq N.$$

Definition: Continuity of a function

We call a function $f : \mathbb{R}^n \to \mathbb{R}^m$ continuous at y if it holds that

$$\lim_{x \to y} f(x) = f(y),$$

i.e. for any sequence $(x_n)_{n \in \mathbb{N}}$ with $\lim_{n\to\infty} x_n = y$ it holds that $\lim_{n\to\infty} f(x_n) = f(y)$.

We call f continuous on $\mathbb{R}^n$ if f is continuous at every $x \in \mathbb{R}^n$.

Lipschitz continuity

Definition: Lipschitz continuity

A function $f : C \subset \mathbb{R}^n \to \mathbb{R}^m$ is called *Lipschitz continuous* with *Lipschitz constant* L if

$$\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$$

holds for all $x, y \in C$.

Examples on the board.

Definition: Local Lipschitz continuity

A function $f : C \subset \mathbb{R}^n \to \mathbb{R}^m$ is called *locally Lipschitz continuous* if for every $x \in C$ there exists $\epsilon > 0$ such that $f_{|B(\epsilon,x)}$ is Lipschitz continuous.

Many of the previous concepts can be generalized to metric or topological spaces, but I stuck to $\mathbb{R}^n$ for the sake of simplicity. Any multivariate calculus book will contain details you might be interested in.

Differentiability

Definition: Partial derivatives

Let $U \subset \mathbb{R}^n$ be open. We call the function $f : U \to \mathbb{R}^m$ partially differentiable at $x \in U$ if

$$\frac{\partial f_i}{\partial x_j}(x) := \lim_{h \to 0} \frac{f_i(x + he_j) - f_i(x)}{h}$$

exists for all $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$, where e_j is the j-th unit normal vector.

If f is partially differentiable at all x and all $\frac{\partial f_i}{\partial x_j} : U \to \mathbb{R}$ are continuous, we call f continuously differentiable, and denote $f \in C^1$.

Jacobi matrix

Definition: Jacobi matrix

For $f : U \subset \mathbb{R}^n \to \mathbb{R}^m \in C^1$, $x \in U$, we call

$$Jf(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & . & . & \frac{\partial f_1}{\partial x_n}(x) \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial f_m}{\partial x_1}(x) & . & . & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

the *Jacobi matrix* of f at x. It is the first derivative of multivariate functions. For $f \in C^1$, Jf itself is a continuous function $Jf : U \to \mathbb{R}^{m \times n}$.

Example on the board $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{2}\|x - y\|^2$.

Composite functions

Chain rule for multivariate functions

Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open. Let

$$f : U \to V \in C^1 \qquad \text{and} \qquad g : V \to \mathbb{R}^k \in C^1.$$

Then the composite function $(g \circ f) : U \to \mathbb{R}^k$ is continuously differentiable and its Jacobian $J(g \circ f)$ is given by

$$J(g \circ f)(x) = (Jg)(f(x)) \cdot Jf(x).$$

Example on the board $g : \mathbb{R}^m \to \mathbb{R}$, $g(x) = \frac{1}{2}\|x - y\|^2$, $f : \mathbb{R}^n \to \mathbb{R}^m$, $f(z) = Az$ for some matrix $A \in \mathbb{R}^{m \times n}$.

Some random important linear algebra definitions

Eigenvalues

Let $A \in \mathbb{R}^{n \times n}$ be a matrix. We say $\lambda \in \mathbb{R}$ is an *eigenvalue* of A if there exists a $v \neq 0$ such that

$$Av = \lambda v.$$

The corresponding v is called an *eigenvector*.

Positive semi-definite matrices

We call $A \in \mathbb{R}^{n \times n}$ *symmetric* if $A = A^T$. A symmetric matrix is called *positive semi-definite* if all its eigenvalues are nonnegative.

Some random important linear algebra definitions

Spectral norm

The spectral norm of a matrix $A \in \mathbb{R}^{n \times m}$ is given by

$$\|A\|_{S^\infty} = \sqrt{\sigma_{max}(A^T A)},$$

where $\sigma_{max}(A^T A)$ is the largest eigenvalue of $A^T A$.

(Generalized) Frobenius norm

For a tensor $A \in \mathbb{R}^{n_1 \times \ldots \times n_r}$ we define

$$\|A\| = \left(\sum_{i_1=1}^{n_1} \ldots \sum_{i_r=1}^{n_r} (A_{i_1,\ldots,i_r})^2 \right)^{1/2}$$

to be the (generalized) Frobenius norm. Notice that it coincides with the ℓ^2 norm for $r = 1$.

Matrix representation of linear operators

We will often deal with linear operators acting on our unknowns. The next theorem will give us an easy way to handle such operators algorithmically:

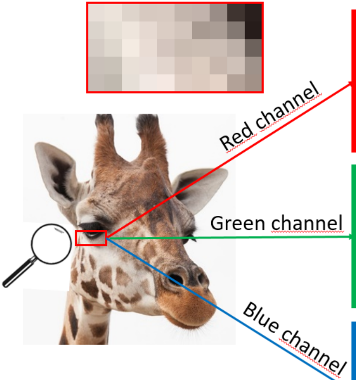
Representation of linear operators as matrices

Let $X = \text{span}(x_1, ..., x_n)$ and $Y = \text{span}(y_1, ..., y_m)$ be arbitrary (final dimensional) vector spaces, and let $\mathcal{A} : X \to Y$ be a linear operator. Then $\mathcal{A}$ admits a representation as a matrix $A \in \mathbb{R}^{m \times n}$ (which depends on the basis $\{x_1, ..., x_n\}$ and $\{y_1, ..., y_m\}$).

To provide an example why this is useful, we will briefly introduce how images are represented digitally.

Representation of images

How does a computer represent an image?



Red channel:

232	236	222	215	216	196	194	154	63	41
219	230	238	230	219	211	168	170	131	38
213	222	244	233	212	202	204	204	189	109
201	201	233	233	226	214	220	201	207	159
197	192	230	241	245	223	221	228	179	162

Green channel:

223	227	212	201	202	182	179	141	52	31
212	223	229	217	205	197	154	157	120	28
204	215	235	220	197	188	190	190	176	98
192	194	224	222	213	197	204	184	193	146
188	185	223	232	232	206	203	210	162	145

Blue channel:

216	218	203	192	193	173	172	135	48	29
204	213	220	209	196	188	145	149	114	26
197	205	226	212	190	177	179	181	168	92
185	184	215	216	205	187	191	174	184	138
181	175	215	227	226	196	191	198	152	135

Optimization with images

An image is represented as a matrix/tensor

$$f \in \mathbb{R}^{n \times m \times c}$$

We can design energies on images. For instance, one might want to smooth an image and penalize its variations, e.g. by defining a discrete gradient operator D,

$$D : \mathbb{R}^{n \times m \times c} \to \mathbb{R}^{n \times m \times c \times 2}$$

$$f \mapsto Df \qquad \text{with } (Df)_{i,j,k,l} = \begin{cases} f_{i,j,k} - f_{i-1,j,k} & \text{if } l = 1 \\ f_{i,j,k} - f_{i,j-1,k} & \text{if } l = 2 \end{cases}$$

and including the squared Frobenius norm $\|Du\|^2$ in the optimization, e.g.

$$E(u) = \|u - f\|^2 + \alpha \|Du\|^2.$$

Vectorization

In cases like our example

$$E(u) = \|u - f\|^2 + \alpha\|Du\|^2$$

it can be very useful to vectorize the problem and turn the linear operator D into a matrix using the previous theorem on representing linear operators!

Advantage 1: We will usually assume $E : \mathbb{R}^n \to \mathbb{R}$, and by looking for the coefficient vector of some basis representation this covers the optimization over any finite dimensional vector space.

Advantage 2: The transpose of a matrix is extremely simple to compute. The analog for an arbitrary linear operator, its *adjoint*, might be much less obvious.

An explicit example on how to represent the discrete gradient as a matrix using the kronecker product will be discussed in the first exercise!

Convexity

Convex energy minimization problems

This lecture is all about

$$\hat{u} \in \arg\min_{u \in C} E(u),$$

where $C \subset \mathbb{R}^n$ convex set, $E : C \to \mathbb{R}$ convex function.

1. What is a convex set?

Definition

A set $C \subset \mathbb{R}^n$ is called convex, if

$$\alpha x + (1-\alpha) y \in C, \quad \forall x, y \in C, \ \forall \alpha \in [0, 1].$$

$\rightarrow$ *Draw a picture.*

$\rightarrow$ *Online TED.*

Convex set - summary of online TED

The following operations preserve the convexity of a set

- Intersection
- Minkowski sum
- Closure
- Interior
- Linear Transformation

The union of convex sets is not convex in general.

Polyhedral sets are always convex, cones are not necessarily convex.

Convex energy minimization problems

Let's get back to what the lecture is all about:

$$\hat{u} \in \arg\min_{u \in C} E(u),$$

where $C \subset \mathbb{R}^n$ convex set, $E : C \to \mathbb{R}$ convex function.

1. What is a convex set? We know this now!

2. What is a convex function?

Definition: Convex Function

We call $E : C \to \mathbb{R}$ a convex function if C is a convex set and for all $u, v \in C$ and all $\theta \in [0, 1]$ it holds that

$$E(\theta u + (1-\theta)v) \leq \theta E(u) + (1-\theta)E(v)$$

We call E strictly convex, if the inequality is strict for all $\theta \in]0, 1[$, and $v \neq u$.

$\rightarrow$ *Draw a picture.*

Convex functions - summary of online TED

The following operations do preserve the convexity of a function

- Summation
- Linear Transformation

The following operations do not preserve the convexity of a function

- Multiplication, Division, Difference
- Composition

The sum of a convex function and a strictly convex function is strictly convex.

Remember to check two conditions to show that a function is convex!

Convex energy minimization problems

Let's get back to what the lecture is all about:

$$\hat{u} \in \arg\min_{u \in C} E(u), \tag{1}$$

where $C \subset \mathbb{R}^n$ convex set, $E : C \to \mathbb{R}$ convex function.

It is sometimes convenient to "introduce" the constraint $u \in C$ into the energy function E itself. We therefore introduce the notion of **extended real valued functions**.

$$E : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}.$$

The minimization problem (1) can then be written as

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} \tilde{E}(u),$$

by defining

$$\tilde{E}(u) = \begin{cases} E(u) & \text{if } u \in C, \\ \infty & \text{else.} \end{cases}$$

Extended real valued functions

Definition

- For $E : \mathbb{R}^n \to \overline{\mathbb{R}}$, we call

$$\text{dom}(E) := \{u \in \mathbb{R}^n \mid E(u) < \infty\}$$

the domain of E.

- We call E proper if $\text{dom}(E) \neq \emptyset$.

Revisiting the definition of convex functions

We call $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ a convex function if

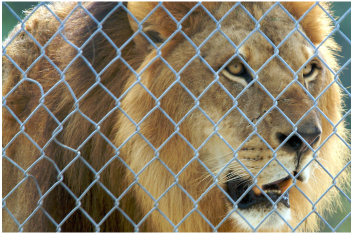
1. $\text{dom}(E)$ is a convex set.
2. For all $u, v \in \text{dom}(E)$ and all $\theta \in [0, 1]$ it holds that

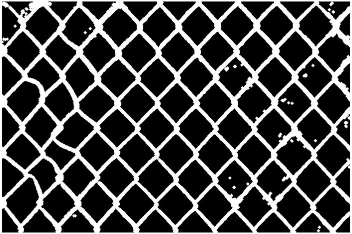
$$E(\theta u + (1-\theta)v) \leq \theta E(u) + (1-\theta)E(v)$$

We call E strictly convex, if the inequality in 2 is strict for all $\theta \in]0, 1[$, and $v \neq u$.

First example of an imaging problem: Inpainting

Example: Inpainting





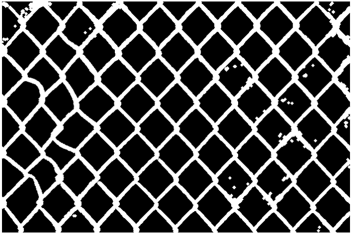
$$\min_{u \in \mathbb{R}^{n \times m}} \sum_{i,j} (u_{i,j} - u_{i-1,j})^2 + (u_{i,j} - u_{i,j-1})^2 \quad \text{s.t. } u_{i,j} = f_{i,j} \; \forall (i,j) \in I$$

with index set I of pixels to keep and suitable boundary conditions.

$\rightarrow$ *Discuss convexity.*

First example of an imaging problem: Inpainting

Example: Inpainting



$$\min_{u \in \mathbb{R}^{n \times m}} \sum_{i,j} (u_{i,j} - u_{i-1,j})^2 + (u_{i,j} - u_{i,j-1})^2 \quad \text{s.t. } u_{i,j} = f_{i,j} \; \forall (i,j) \in I$$

with index set I of pixels to keep and suitable boundary conditions.

$\rightarrow$ *Discuss convexity.*

Epigraph of a function

Is there a connection between convex sets and functions?

Defintion: Epigraph

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function mapping into the extended real line. Then

$$\text{epi}(E) := \{(u, \alpha) \mid E(u) \leq \alpha\}$$

is called the *epigraph* of the function E.

$\rightarrow$ *Draw a picture.*

Theorem

A proper function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if it's epigraph is convex

Proof: Second exercise sheet.

Properties of convex functions

What is so special about convex functions?

Theorem

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Any local minimum of E is global.

Proof: Board.

Theorem: Monotonicity of the gradient

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, convex and differentiable at $u \in \text{dom}(E)$.

$$E(v) - E(u) - \langle \nabla E(u), v - u \rangle \geq 0 \qquad \forall v \in \mathbb{R}^n$$

Proof: Later.

Conclusion

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, convex and differentiable at $u \in \text{dom}(E)$. If $\nabla E(u) = 0$ then u is a global minimum of E.

Illustration on the board.

Convex energy minimization problems

We said this lecture is all about

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} E(u),$$

where $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function.

Does a minimizer of such a function even exist?

Existence of minimizers

Defintion: Lower semi-continuity (l.s.c.)

We call the function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ lower semi-continuous (l.s.c.), if for all u it holds that

$$\liminf_{v \to u} E(v) \geq E(u).$$

Theorem: Existence of minimizers

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be l.s.c. and let there exist an α such that the sublevelset

$$\{u \in \mathbb{R}^n \mid E(u) \leq \alpha\}$$

is nonempty and bounded, then

$$\hat{u} \in \arg\min_u E(u)$$

exists.

Proof: Board.

Closedness and lower semi-continuity

Defintion: Closed function

We call the function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ closed if it's epigraph is closed.

Theorem: Equivalence of l.s.c. and closedness

For $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ the following two statements are equivalent

- E is lower semi-continuous (l.s.c.).
- E is closed.

Proof: Board.

OnlineTED!

Continuity of convex functions

Continuity of Convex Functions

If $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex, then E is locally Lipschitz (and hence continuous) on int(dom(E)).

Proof in 1d: Exercise for yourself (solution will be online)

$\to$ *Board: Considering the interior is important!*

Conclusion

If $E : \mathbb{R}^n \to \mathbb{R}$ is convex, then E is continuous.

Existence of minimizers part II

Definition: Coercivity

A function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called *coercive* if $E(v_n) \to \infty$ for all sequences $(v_n)_n$ with $\|v_n\| \to \infty$.

Remark: Coercivity implies that there exists a bounded sublevelset.

Existence of a minimizer for function with full domain

Let $E : \mathbb{R}^n \to \mathbb{R}$ be convex and coercive, then an element $\hat{u} \in \arg\min_u E(u)$ exists.

Proof:

- $\mathrm{dom}(E) = \mathbb{R}^n$, E convex $\Rightarrow$ E is continuous.
- E is coercive, i.e. there exists a non-empty bounded sublevelset.

Uniqueness

When is

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} E(u)$$

unique?

Theorem: Uniqueness

If $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strictly convex, then there exists at most one local minimum which is the unique global minimum.

Proof: Simple computation.

Optimality conditions

How can we determine if

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} E(u)? \tag{1}$$

In other words,

what is the optimality condition for (1)?

Consider a differentiable E and remember analysis I:

Necessary condition for local extremum is

$$\nabla E(\hat{u}) = 0$$

Sufficient condition? **Convexity!**

Examples: Derivatives of convex functions

What are the optimality conditions for ...

... $E(u) = \|u - f\|_2^2 = \sum_{i=1}^n (u_i - f_i)^2$?

... $E(u) = \|Au - f\|_2^2$ for a matrix $A \in \mathbb{R}^{m \times n}$?

... $E(u) = \|u\|_1 = \sum_{i=1}^n |u_i|$?

We need a theory for non-differentiable functions!

Illustrate ℓ^1 case for discussion.

The subdifferential

Definition: Subdifferential

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. We call

$$\partial E(u) = \{p \in \mathbb{R}^n \mid E(v) - E(u) - \langle p, v - u \rangle \geq 0, \ \forall v \in \mathbb{R}^n\}$$

the subdifferential of E at u.

- Elements of $\partial E(u)$ are called subgradients.
- If $\partial E(u) \neq \emptyset$, we call E subdifferentiable at u.
- By convention, $\partial E(u) = \emptyset$ for $u \notin \text{dom}(E)$.

Theorem: Optimality condition

Let $0 \in \partial E(\hat{u})$. Then $\hat{u} \in \arg\min_u E(u)$.

The subdifferential

Example functions for discussing the subdifferential:

- The absolute value function.
- Functional
$$E(u) = \begin{cases} 0 & \text{if } u \geq 0 \\ \infty & \text{else.} \end{cases}$$
- $E(u) = \frac{1}{2}\|u\|^2$

Subdifferential and derivatives

Let the convex function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be differentiable at $u \in \text{int}(\text{dom}(E))$. Then

$$\partial E(u) = \{\nabla E(u)\}.$$

Proof: Exercise.

This also proves the "Theorem: Monotonicity of the gradient".

The subdifferential

Geometric interpretation of subgradients:

Any subgradient $p \in \partial E(u)$ represents a non-vertical supporting hyperplane to $\text{epi}(E)$ at $(u, E(u))$.

Definition

A supporting hyperplane to a set $S \subset \mathbb{R}^n$ is a hyperplane $\{x \in \mathbb{R}^n \mid \langle a, x \rangle = b\}$, $a \neq 0$, such that

- $S \subset \{x \in \mathbb{R}^n \mid \langle a, x \rangle \leq b\}$ or $S \subset \{x \in \mathbb{R}^n \mid \langle a, x \rangle \geq b\}$
- $\exists y \in \partial S$ (the boundary of S) such that $\langle a, y \rangle = b$.

Let $p \in \partial E(u)$. Then

$$\begin{aligned}
& E(v) - E(u) - \langle p, v-u \rangle \geq 0 && \forall v \in \mathbb{R}^n \\
\Rightarrow \quad & \alpha - E(u) - \langle p, v-u \rangle \geq 0 && \forall (v, \alpha) \in \text{epi}(E) \\
\Rightarrow \quad & \left\langle \begin{bmatrix} -p \\ 1 \end{bmatrix}, \begin{bmatrix} v \\ \alpha \end{bmatrix} - \begin{bmatrix} u \\ E(u) \end{bmatrix} \right\rangle \geq 0 && \forall (v, \alpha) \in \text{epi}(E).
\end{aligned}$$

$\rightarrow$ *Draw image on the board.*

The subdifferential

Is any convex E subdifferentiable at $x \in \text{dom}(E)$?

$$E(u) = \begin{cases} -\sqrt{u} & \text{if } u \geq 0 \\ \infty & \text{else.} \end{cases}$$

Definition: Relative Interior

The *relative interior* of a convex set M is defined as

$$\text{ri}(M) := \{x \in M \mid \forall y \in M,\ \exists \lambda > 1,\ \text{s.t. } \lambda x + (1-\lambda) y \in M\}$$

Theorem: Subdifferentiability[a]

[a] Rockafellar, Convex Analysis, Theorem 23.4

If E is a proper convex function and $u \in \text{ri}(\text{dom}(E))$, then $\partial E(u)$ is non-empty. The set $\partial E(u)$ is non-empty and bounded if and only if $u \in \text{int}(\text{dom}(E))$.

Partial proof on the board, full proof Rockafellar.

Subdifferential rules

Theorem: Sum rule[a]

[a] Rockafellar, Convex Analysis, Theorem 23.8

Let E_1, E_2 be convex functions such that

$$\text{ri}(\text{dom}(E_1)) \cap \text{ri}(\text{dom}(E_2)) \neq \emptyset,$$

then it holds that

$$\partial(E_1 + E_2)(u) = \partial E_1(u) + \partial E_2(u).$$

Example: Minimize $(u - f)^2 + \iota_{u \geq 0}(u)$.

Example: Minimize $0.5(u - f)^2 + \alpha|u|$.

Example: Minimize $0.5\|u - f\|_2^2 + \alpha|\langle u, a\rangle|$ for some $a \in \mathbb{R}^n$.

Example: Minimize $0.5\|u - f\|_2^2 + \alpha\|Wu\|_1$ for $W \in \mathbb{R}^{n \times n}$ being orthonormal.

Subdifferential rules

Theorem: Chain rule[a]

[a] Rockafellar, Convex Analysis, Theorem 23.9

If $A \in \mathbb{R}^{m \times n}$, $E : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is convex, and $\text{ri}(\text{dom}(E)) \cap \text{range}(A) \neq \emptyset$, then

$$\partial(E \circ A)(u) = A^* \partial E(Au)$$

Example: Minimize $\|Au - f\|_2^2$.

Example: Minimize $\|D_1 u - f\|_2^2 + \alpha \|D_2 u + g\|_1$, for D_1 and D_2 being diagonal, (and D_2 being invertible).

Summary

- **Convex functions**
  - Every local minimum is global
  - First order optimality condition is sufficient
- The **optimality condition** for $\hat{u}$ to minimize E is

$$0 \in \partial E(\hat{u})$$

- **The subdifferential** $\partial E(u)$
  - is set valued.
  - generalizes the derivative.
  - $\partial E(u) = \{\nabla E(u)\}$ is E is differentiable at u.
  - can be identified with supporting hyperplanes to epi(E).
  - Obeys the "usual" sum and chain rules.

We now have all tools that are necessary to discuss a first class of minimization algorithms for determining

$$\hat{u} \in \underset{u}{\operatorname{argmin}}\, E(u)$$

Up next: Implementation, convergence, applications!