1-Forms

$\underline{a}$, $\underline{u}$, $\underline{v}$, $\underline{x}$	real 1-form in $\mathbb{R}^n$ with i^{th} entry a_i, etc.; computed as a $1 \times n$ row matrix $\underline{a} = [a_1\ a_2\ \cdots\ a_n]$
$\underline{0}$, $\underline{1}$	real 1-forms with constant entries
$\underline{z}(M)$	"mean" 1-form of a matrix, where entry z_j is the mean of column j of M
$\underline{P}$, $\underline{Q}$	Probability distributions

General Matrices

A, C, M, X, Y, Z	matrix; entries are a_{ij}, etc.
B	symmetric positive definite matrix
D, K	symmetric matrix, positive definite or positive semidefinite
R	upper-triangular matrix
Q, U	orthogonal matrix
V	orthogonal matrix; may give eigenvectors of another matrix
W	*context dependent*; general matrix or symmetric positive semidefinite matrix

Special Functions

$H(u)$	Heaviside step function; $H(v) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if} \quad v < 0 \\ 1 & \text{if} \quad v \geq 0 \end{cases}$

Greek Symbols

α, α_i, $\vec{\alpha}$	real number, *context dependent*; typically a known value or, in SVM, Lagrange multiplier
β, γ, ρ	real numbers, especially as known values
η	positive real number; typically, $0 < \eta < 1$ as a learning rate
λ, λ_i	real number, *context dependent*; eigenvalue or Lagrange multiplier or regularizer
μ, μ_i	real number, *context dependent*; free parameter or Lagrange multiplier or mean value
ν	real number; generally, variable parameter
ϕ, θ, ψ	real number
$\vec{\rho}$	real vector, possibly indexed
Λ	diagonal matrix where λ_i is an eigenvalue of another matrix
Σ	diagonal matrix, or diagonal-like matrix such as the singular-value matrix

Differentials

f', $\dfrac{df}{dt}$	derivative of $f:\mathbb{R} \to \mathbb{R}$
f'', $\dfrac{d^2f}{dt^2}$	derivative of $f':\mathbb{R} \to \mathbb{R}$
$\dfrac{\partial f}{\partial w_i}(\vec{w})$	partial derivative of $\vec{f}:\mathbb{R}^n \to \mathbb{R}$ with respect to w_i
$\underline{\nabla}\, f(\vec{w})$	1-form derivative of $f:\mathbb{R}^n \to \mathbb{R}$; i^{th} entry is $\dfrac{\partial f}{\partial w_i}$
$\dfrac{\partial^2 f}{\partial w_i \partial w_j}(\vec{w})$	partial derivative of $\dfrac{\partial f}{\partial w_j}:\mathbb{R}^n \to \mathbb{R}$ with respect to w_i
$D_{\vec{v}} f(\vec{w})$	directional derivative of $f:\mathbb{R}^n \to \mathbb{R}$ at $\vec{w}$ in direction $\vec{v}$
$J_{\vec{f}}(\vec{w}_0)$	Jacobian matrix of $\vec{f}:\mathbb{R}^n \to \mathbb{R}^m$, evaluated at $\vec{w}_0$; entry (i,j) is $\dfrac{\partial f_i}{\partial w_j}(\vec{w}_0)$ so the i^{th} row of $J_{\vec{f}}(\vec{w}_0)$ is $\underline{\nabla}\, f_i(\vec{w}_0)$
$\nabla^2 f(\vec{w}_0)$	Hessian matrix of $f:\mathbb{R}^n \to \mathbb{R}$, evaluated at $\vec{w}_0$; entry (i,j) is $\dfrac{\partial^2 f}{\partial w_i \partial w_j}(\vec{w}_0)$

Sets

$\mathbb{N}$	set of natural numbers, including zero
$\mathbb{N}_+$	set of positive natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	set of real vectors, each vector having n entries
$\mathbb{R}^n_+$	set of real vectors where each entry of each vector is non-negative
$\mathbb{R}^n_{++}$	set of real vectors where each entry of each vector is positive
$\mathbb{S}_C(f,l)$	level curve of function f at level l
$\mathbb{S}_L(f,l)$	level set of function f at level l
$\mathbb{V}$	vector space over the field of real numbers
$\mathbb{A}$	events a_i
$\mathbb{A}(M,\vec{c})$	Affine space; set of vectors $\vec{u}$ that are solutions to $M\vec{w} = \vec{c}$
$\mathbb{F}$	Feasible solutions of a constrained optimization problem

CISC 371 Class 1

Introduction to Optimization

Texts: [1] pp.1–10

Main Concepts:
- *Optimum: maximum or minimum*
- *Unconstrained optimization*
- *constrained optimization*

Sample Problem, Spatial Localization: Which point is at the "middle" of a set of given points?

In this course, optimization is the process of selecting a "best" member of a set according to a criterion. We will explore sets that may be vector spaces, vector subspaces, or subsets of vector subspaces. For brevity, we will refer to any member of a set $\mathbb{V}$ as a *point*, which we will write as t or, if $\mathbb{V}$ is a vector space, as $\vec{w}$. In an axiomatic Euclidean geometry, a point is the simplest possible geometrical object and, for this course, a vector satisfies this definition.

The criterion that we will use is a function, which in this course is a mapping from a set of points to the real numbers. The proper usage would be to use a "functional", which is a map from a vector space to the scalar of the vector space; because current usage in optimization is the word "function", we use this term here despite its ambiguity.

A simple example of optimization is a function that maps a real number to a real number, which can be written as

$$f : \mathbb{R} \to \mathbb{R}$$

A simpler, more familiar notation for one such function might be

$$f_1(t) \stackrel{\text{def}}{=} \frac{-t}{t^2+1} \tag{1.1}$$

In this course, we will often refer to such a function as an *objective function*. The term "objective" will be used as an abbreviation for the longer term. A point, from the domain of f, is the *argument* of the function.

An example of the objective function of Equation 1.1 is plotted in Figure 1.1. We see that there are two "best" values. From visual examination of the plot, we can see that the maximum value of f_1 is $+0.5$ and the minimum value of f_1 is -0.5; these occur at the point $t = +1$ and at the