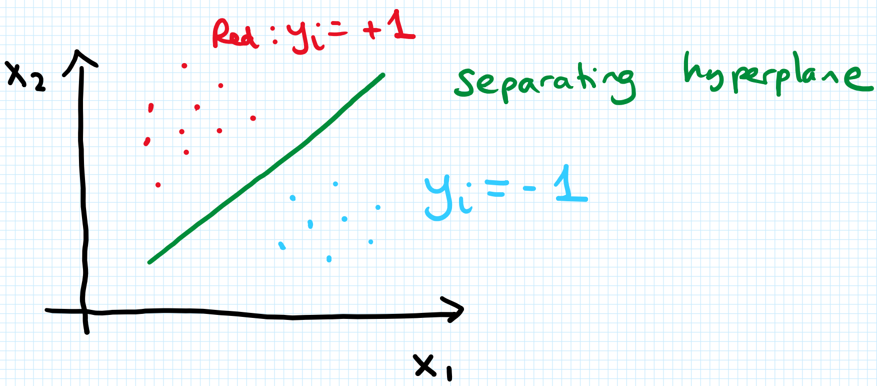
Second part of course: Optimization with constraints

- Example of unconstrained optimization: ANNs
- Example of constrained optimization: SVMs

Binary classification via SVMs



The best separating hyperplane is found by solving an OP:

$$\min_{\underline{w} \in \mathbb{R}^n} \frac{1}{2}\|\underline{w}\|_2^2 \quad \text{subject to} \quad \underbrace{\begin{matrix} y_i(\underline{x}_i \cdot \underline{w} + b) \geq 0 \\ i = 1, 2, \dots, m \end{matrix}}_{\text{Constraints}}$$

Solved using the KKT conditions.

Aim of this part of the course: A proof from first principles of the KKT conditions (Nocedal and Wright).

KKT: 40s, 50s, 60s

Dual problem: Back to Hamilton...

Dual problem : Back to Hamilton.

Karush–Kuhn–Tucker conditions

17 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

In mathematical optimization, the **Karush–Kuhn–Tucker** (**KKT**) **conditions**, also known as the **Kuhn–Tucker conditions**, are first derivative tests (sometimes called first-order necessary conditions) for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied.

Allowing inequality constraints, the KKT approach to nonlinear programming generalizes the method of Lagrange multipliers, which allows only equality constraints. Similar to the Lagrange approach, the constrained maximization (minimization) problem is rewritten as a Lagrange function whose optimal point is a global maximum or minimum over the domain of the choice variables and a global minimum (maximum) over the multipliers. The Karush–Kuhn–Tucker theorem is sometimes referred to as the saddle-point theorem.[1]

The KKT conditions were originally named after Harold W. Kuhn and Albert W. Tucker, who first published the conditions in 1951.[2] Later scholars discovered that the necessary conditions for this problem had been stated in an unpublished master's thesis by William Karush in 1939.[3][4]

Problem statement for this part of the course (Ch. 12)

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \quad \text{subject to} \quad \begin{cases} c_i(\underline{x}) = 0 & i \in \mathcal{E} \\ c_i(\underline{x}) \geq 0 & i \in \mathcal{I} \end{cases}$$

Here, $\mathcal{E}$ is a set of indices labelling equality constraints and $\mathcal{I}$ is a set of indices labelling inequality constraints.

We denote a solution to the OP as $\underline{x}^*$.

Feasible region:

$$\Omega = \left\{ \underline{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} c_i(\underline{x}) = 0 & i \in \mathcal{E} \\ c_i(\underline{x}) \geq 0 & i \in \mathcal{I} \end{array} \right\}$$

Recall the necessary conditions for optimality for the unconstrained problem:

- $\nabla f = 0$ @ $\underline{x} = \underline{x}_*$
- $B_{ij} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{\underline{x}_*}$ is positive-semi-definite.

Conditions for the constrained problem are more complicated, but we will derive them anyway.

Learning strategy:

- Motivating examples
- Definitions
- Rigorous proofs

Second exam

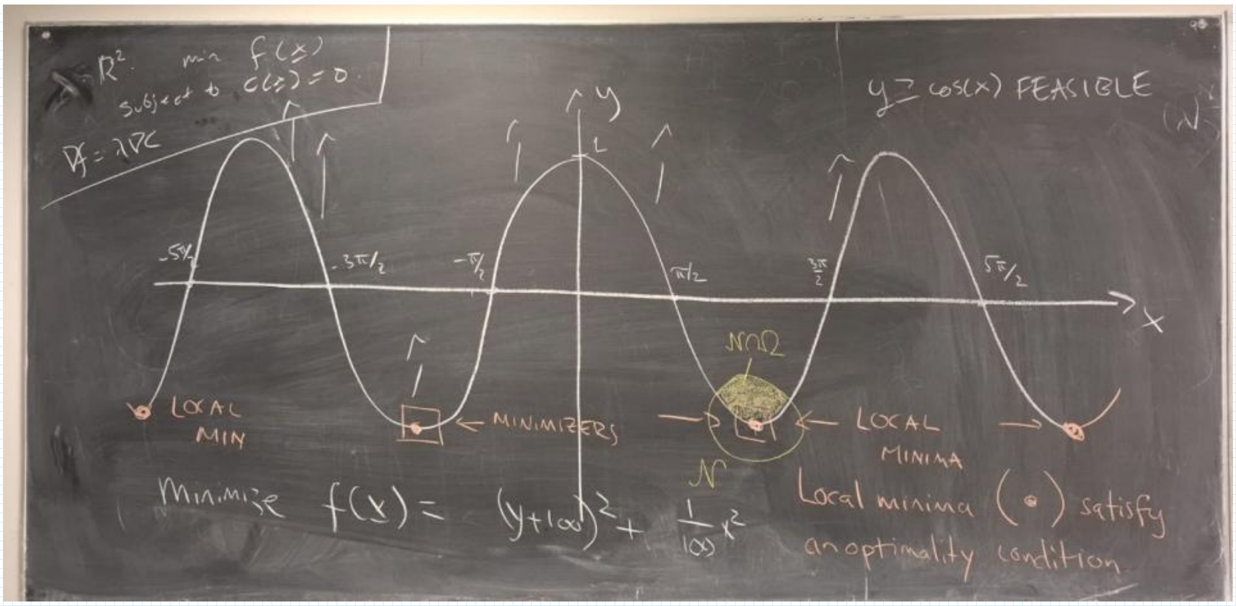
- Exercises
- List of proofs

Examples (§ 12.1.1)

$$\min_{\underline{x} \in \mathbb{R}^2} f(\underline{x}) = (y+100)^2 + \frac{1}{100} x^2$$

subject to: $y - \cos(x) \geq 0$.

Feasible region: $y \geq \cos(x)$



Solution to the OP is to make y as negative as possible within a neighbourhood, giving $y = -1$ and $x = k\pi$, $k = \pm 1, \pm 3, \dots$

This gives infinitely many <u>local minimizers</u>.

This is a strange example where the unconstrained problem has a single global minimizer whereas the constrained problem has <u>infinitely</u> many local minimizers. We make <u>this more precise</u> in the <u>exercises</u>.

The corresponding unconstrained case:

$$x_* = 0, \qquad y_* = -100.$$

$x_* = 0, \quad y_* = -100.$

$$\min f(\underline{x}) = 0 \qquad \text{unconstrained}$$

$$\min f(\underline{x}) = 99^2 + \frac{1}{100} k^2 \pi^2$$

$$\min_{\underline{x} \in \mathbb{R}^2} f(\underline{x}) \leq \min_{\underline{x} \in \Omega} f(\underline{x})$$

This is a general result: Since $\Omega \subset \mathbb{R}^2$,

$$\min_{\underline{x} \in \mathbb{R}^2} f(\underline{x}) \leq \min_{\underline{x} \in \Omega} f(\underline{x})$$

Once you add some constraints, you don't necessarily pick out the global min.

For a non-PC application, ask me after class!

Introduce some definitions

- $\underline{x}_*$ is a local sl^n of the OP if
  - $\underline{x}_* \in \Omega$
  - There exists a neighbourhood $\underline{x}_* \ni \mathcal{N}$ such that:

$$f(\underline{x}_*) \leq f(\underline{x}) \;\; \forall \;\; \underline{x} \in \mathcal{N} \cap \Omega$$

- $\underline{x}_*$ is a strict local s|^n of the OP if
  - $\underline{x}_* \in \Omega$
  - There exists a neighbourhood $\underline{x}_* \ni \mathcal{N}$ such that

$$f(\underline{x}_*) < f(\underline{x}) \quad \forall \; \underline{x} \neq \underline{x}_* \in \mathcal{N} \cap \Omega$$

- $\underline{x}_*$ is an isolated local s|^n of the OP if
  - $\underline{x}_* \in \Omega$
  - There exists a neighbourhood $\underline{x}_* \ni \mathcal{N}$ such that $\underline{x}_*$ is the only minimizer in $\mathcal{N} \cap \Omega$.

<u>Smoothness</u> : § 12.1.2

- We need the cost function to be differentiable.
- The boundary of the feasible region can be:
  - differentiable
  - piecewise-differentiable

<u>Example</u> : $\Omega = \{ \underline{x} \in \mathbb{R}^2 \mid \|\underline{x}\|_1 \leq 1 \}$

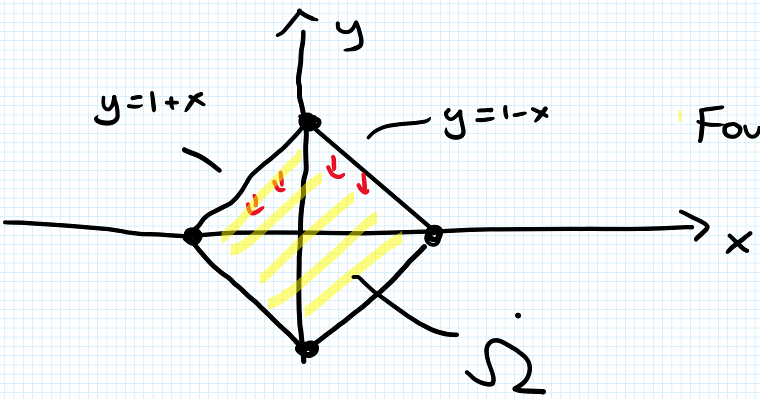
<u>OR</u> $\Omega = \{ \underline{x} \in \mathbb{R}^2 \mid |x| + |y| \leq 1 \}$

The boundary $\partial\Omega$ can be broken up into four line

The boundary $\partial\Omega$ can be broken up into four line segments:

1^{st} Quadrant: $x + y \leq 1 \Rightarrow y \leq 1 - x$

2^{nd} Quadrant: $-x + y \leq 1 \Rightarrow y \leq 1 + x$



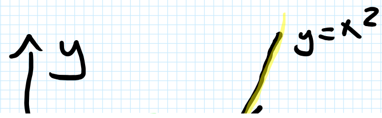
Four cusps (•)

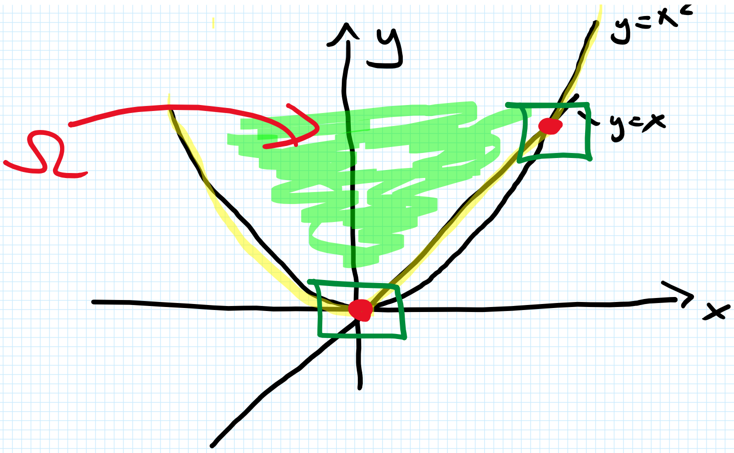
Piecewise-smooth boundary $\partial\Omega$

Sometimes we can accept a non-smooth cost function but we have to do a trick:

- non-smooth cost function for an unconstrained OP
- increase the dimension of the OP
- Turn the problem into a constrained OP
- Make sure ~~the~~ any non-smooth points are on $\partial\Omega$.

Example: $f(x) = \max(x^2, x)$ in 1D





OP1:

$\min f(x), \; x \in \mathbb{R}$

$$\Omega = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} y \geq x \\ y \geq x^2 \end{array} \right\}$$

OP2: $\tilde{f} : \mathbb{R}^2 \to \mathbb{R}, \; (x,y) \to y$

$\min_{\underline{x} \in \mathbb{R}^2} \tilde{f}(\underline{x})$ subject to $\underline{x} \in \Omega$.

Application: Ridge regression:

$$f(\underline{x}) = \underbrace{\mathcal{L}(\underline{x})}_{\text{Loss function}} + \underbrace{\eta \|\underline{x}\|_1}_{\text{Penalty, penalizing over-fitting}}$$

Penalty term $\|\underline{x}\|_1$ not differentiable, but we can turn this into a constrained OP and move the constraints to the boundary (Exercises).

Plan for today:

- Terminology
- Example – 1 equality constraint
- Example – 1 inequality constraint
- Draw general conclusions

OP (constrained optimization)

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \quad \text{subject to:} \quad \begin{cases} c_i(\underline{x}) = 0, & i \in \mathcal{E} \\ c_i(\underline{x}) \geq 0, & i \in \mathcal{I} \end{cases}$$

$$\Omega = \left\{ \underline{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} c_i(\underline{x}) = 0, \ i \in \mathcal{E} \\ c_i(\underline{x}) \geq 0, \ i \in \mathcal{I} \end{array} \right\}$$ FEASIBLE REGION

We use $\underline{x}_*$ to denote a feasible point, not necessarily optimal.

Active set (§ 12.1.3)

<u>Def</u>: The active set $\mathcal{A}(\underline{x})$ at any feasible point $\underline{x} \in \Omega$ is a set of indices:

- The indices of all the equality constraints
- The indices of the inequality constraints satisfying $c_i(\underline{x}) = 0$

<u>Or</u>:

$$\mathcal{A}(\underline{x}) = \mathcal{E} \cup \{ i \in \mathcal{I} \mid c_i(\underline{x}) = 0 \}$$

Terminology: We have the notion of an active constraint.

An inequality constraint $c_i(\underline{x}) \geq 0$ is active at $\underline{x}$

An inequality constraint $c_i(\underline{x}) \geq 0$ is active at $\underline{x}$ if $c_i(\underline{x}) = 0$. The constraint is inactive if $c_i(\underline{x}) > 0$.

§ 12.2 — Worked Example — A single equality constraint.

OP: Minimize $f(\underline{x}) = x + y$ ($\underline{x} \in \mathbb{R}^2$) subject to $c_1(\underline{x}) = 0$, where $c_1(\underline{x}) = 2 - x^2 - y^2$

<u>Remark</u>: $\mathcal{I} = \emptyset$, $\mathcal{E} = \{1\}$

<u>Analytical solution</u>: $c_1(\underline{x}) = 0 \Rightarrow x^2 + y^2 = 2$

Hence, we are constrained to be on a circle of radius $\sqrt{2}$. Re-parametrize the OP as an unconstrained problem on a lower-dimensional sub-manifold:

$$x = \sqrt{2}\cos\theta, \qquad y = \sqrt{2}\sin\theta$$

$$f(\underline{x}) = \sqrt{2}\left(\cos\theta + \sin\theta\right) = \tilde{f}(\theta)$$

OP on sub-manifold:

minimize $\tilde{f}(\theta)$ (unconstrained)

$$\frac{d\tilde{f}}{d\theta} = \sqrt{2}\left(-\sin\theta + \cos\theta\right)$$

$\theta = \pi/$

$$\frac{d\tilde{f}}{d\theta}$$

$$\frac{d\tilde{f}}{d\theta} = 0 \Rightarrow \sin\theta = \cos\theta \Rightarrow \begin{cases} \theta = \pi/4 \\ \theta = 5\pi/4 \end{cases}$$

$$\tilde{f}(\theta = \pi/4) = \sqrt{2}\left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\right) = 2 \qquad \text{MAX}$$

$$\tilde{f}\left(\theta = \frac{5\pi}{4}\right) = \sqrt{2}\left(-\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\right) = -2 \qquad \text{MIN}$$

Back to $\underline{x} = (x,y) = (\sqrt{2}\cos\theta, \sqrt{2}\sin\theta)$

$$\boxed{\underline{x}_* = (-1,-1)}$$

$$\nabla f = (1,1)$$

$$\nabla c_1 = (-2x, -2y)$$

$$\nabla c_1(\underline{x}_*) = (2,2)$$

$$\Rightarrow \quad \nabla f(\underline{x}_*) \,\|\, \nabla c_1(\underline{x}_*)$$

$\Rightarrow$ there exists a scalar λ_1^* $(= 1/2)$ such that

$$\nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*)$$

Hence, λ_1^* is the Lagrange multiplier.

General case — a single equality constraint (§12.3)

$\underline{\text{OP}}$: $\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x})$, subject to $c_1(\underline{x}) = 0$.

We $\underline{\text{show}}$ that at the solution to the OP ($= \underline{x}_*$), there exists a scalar λ_1^* positive or negative, such that

$$\nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*)$$

$\underline{\text{Idea}}$: Take a feasible point $\underline{x}$, not necessarily optimal. Take a neighbouring feasible point $\underline{x} + \underline{\delta}$:

$$c_1(\underline{x} + \underline{\delta}) = 0$$

$$\Rightarrow \cancel{c_1(\underline{x})} + \underline{\delta} \cdot \underline{\nabla} c_1(\underline{x}) = 0, \quad \|\underline{\delta}\|_2 \text{ sufficiently small}$$

$$\Rightarrow \underline{\delta} \cdot \underline{\nabla} c_1(\underline{x}) = 0$$

Apply this to the s^{n} of the OP ($= \underline{x}_*$):

$$\boxed{\underline{\delta} \cdot \underline{\nabla} c_1(\underline{x}_*) = 0}$$

Since $\underline{x}_*$ is optimal:

$$f(\underline{x}_* + \underline{\delta}) \geq f(\underline{x}_*)$$

Taylor expansion:

$$\cancel{f(\underline{x}_*)} + \underline{\delta} \cdot \underline{\nabla} f(\underline{x}_*) \geq \cancel{f(\underline{x}_*)}$$

$$\Rightarrow \quad \underline{\delta} \cdot \nabla f(\underline{x}_*) \geq 0$$

Go through the same argument, with $\underline{\delta}$ replaced with $-\underline{\delta}$ ($\underline{x}_* \pm \underline{\delta}$ are feasible):

$$\Rightarrow \quad -\underline{\delta} \cdot \nabla f(\underline{x}_*) \geq 0$$

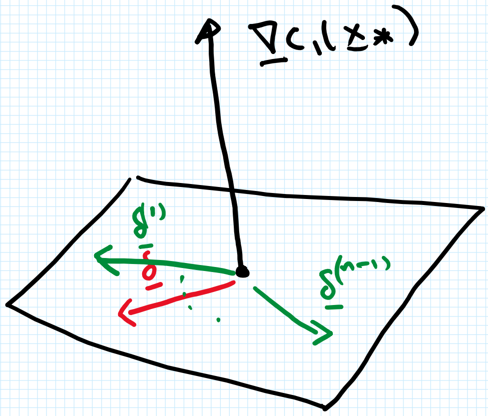
$$\Rightarrow \quad \boxed{\underline{\delta} \cdot \nabla f(\underline{x}_*) = 0}$$

Summarize:

$$\boxed{\underline{\delta} \cdot \nabla f(\underline{x}_*) = 0} \qquad (1)$$

$$\underline{\delta} \cdot \nabla c_1(\underline{x}_*) = 0 \qquad (2)$$

By (2) $\underline{\delta}$ is a hyperplane $\perp^{ar}$ to $\nabla c_1(\underline{x}_*)$:



The hyperplane $\underline{\delta}$ is spanned by $(n-1)$ linearly independent vectors $\underline{\delta}^{(1)}, \dots, \underline{\delta}^{(n-1)}$.

Back to Eq (1):

Back to Eq (1):

$$\underline{\delta}^{(i)} \cdot \nabla f(\underline{x}_*) = 0 \quad , \quad i \in \{1, 2, \ldots, n-1\}$$

Hence, $\nabla f(\underline{x}_*)$ must point normal to the hyperplane, so there exists a constant λ_1^* positive or negative such that

$$\boxed{\nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*) .}$$

Projection operator (§ 12.4)

- $\underline{\delta}$ is like the search direction in SD
- But with constraints we can't search in any direction
- The projection operator tells us which directions we can search in.

$$\mathbb{P} = \mathbb{I} - \frac{\nabla c_1 \otimes \nabla c_1}{\|\nabla c_1\|_2^{\,2}}$$

PROJECTION OPERATOR, SINGLE EQUALITY CONSTRAINT

In components:

$$P_{ij} = \delta_{ij} - \frac{1}{\|\nabla c_1\|_2^2} \frac{\partial c_1}{\partial x_i} \frac{\partial c_1}{\partial x_j} \quad \Bigg\} \quad \underline{\nabla} c_1 = \hat{n} \, \|\underline{\nabla} c_1\|_2$$

For any vector $\underline{v}$, $\mathbb{P}\underline{v} \perp^{r} \nabla c_1$

Proof :

$$\langle \mathbb{P}\underline{v}, \nabla c_1 \rangle = \sum_i (\mathbb{P}\underline{v})_i \frac{\partial c_1}{\partial x_i}$$

$$= \sum_{i,j} P_{ij} v_j \frac{\partial c_1}{\partial x_i}$$

$$= \sum_{ij} (\delta_{ij} - \hat{m}_i \hat{m}_j) v_j \hat{m}_i \|\nabla c_1\|_2$$

$$= \|\nabla c_1\|_2 \sum_{ij} \left(\delta_{ij} v_j \hat{m}_i - \hat{m}_i \hat{m}_j v_j \hat{m}_i \right)$$

$$= \|\nabla c_1\|_2 \left[\underline{v} \cdot \hat{\underline{m}} - \underbrace{(\hat{m} \cdot \hat{m})}_{=1} \underline{v} \cdot \hat{m} \right]$$

$$= 0$$

$\therefore \mathbb{P}\underline{v} \perp^{\underline{ar}} \nabla c_1$ ∎

So, if we were to do steepest descents for a single equality constraint, we would go out in a direction

$$\hat{\underline{n}} = \frac{\underline{s}}{\|\underline{s}\|_2}, \qquad \underline{s} = -\mathbb{P}\nabla f$$

For a sequence of iterates $\{\underline{x}_k\}_{k=0}^{\infty}$ with

$$\underline{x}_{k+1} = \underline{x}_k + \alpha \hat{n}$$

$$f(\underline{x}_{k+1}) = f(\underline{x}_k + \alpha \hat{n})$$

$$= f(\underline{x}_k) + \alpha\, \hat{n} \cdot \nabla f(\underline{x}_k), \quad \alpha \text{ suff. small}$$

Reduction:

$$\Delta_k = f(\underline{x}_{k+1}) - f(\underline{x}_k)$$

$$= \alpha\, \hat{n} \cdot \nabla f(\underline{x}_k)$$

$$= \alpha\, \frac{\underline{s}}{\|\underline{s}\|_2} \cdot \nabla f(\underline{x}_k)$$

$$= \frac{\alpha}{\|\underline{s}\|_2} \left[- \mathbb{P} \nabla f(\underline{x}_k) \right] \cdot \nabla f(\underline{x}_k)$$

$$\propto - \langle \mathbb{P} \underbrace{\nabla f(\underline{x}_k)}_{\underline{g}}, \nabla f(\underline{x}_k) \rangle$$

$$= - \sum_{i,j} (P_{ij}\, g_j)\, g_i$$

$$= - \sum_{ij} (\delta_{ij} - \hat{m}_i \hat{m}_j)\, g_i\, g_j$$

$$= - \left[\|g\|_2^2 - (\hat{m} \cdot g)^2 \right]$$

$$= - \left[\|g\|_2^2 - \|g\|_2^2 \cos^2\theta \right]$$

$$= - \|g\|_2^2 \underbrace{(1 - \cos^2\theta)}_{\geq 0}$$

$$= -\|y\|_2 \, (1 - \cos\theta)$$

$$\underbrace{\qquad\qquad \geq 0 \qquad}_{\leq 0}$$

$\Rightarrow \Delta_k \leq 0$, for α suff. small.

Summarizing, for a single equality constraint, we can introduce the Lagrangian

$$\mathcal{L}(\underline{x}, \lambda) = f(\underline{x}) - \lambda c_1(\underline{x})$$

$$\nabla_x \mathcal{L} = \nabla f(\underline{x}) - \lambda \nabla c_1(\underline{x})$$

At the minimum $\underline{x}_*$ we have found, there exists a scalar λ_1^* such that

$$\left\{ \begin{array}{l} \nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*) \\ c_1(\underline{x}_*) = 0 \end{array} \right\}$$

I.e. At the minimum $(\underline{x} = \underline{x}_*, \lambda = \lambda_1^*)$

$$\left\{ \begin{array}{l} \nabla_x \mathcal{L}(\underline{x}_*, \lambda_1^*) = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda}(\underline{x}_*, \lambda_1^*) = 0 \end{array} \right\}$$

I.e. the first-order optimality condition for an unconstrained problem in a higherfor an unconstrained problem in a higher-dimensional space $(\underline{x}, \lambda)$.

A single inequality constraint (§ 13.2)

OP:

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \quad \text{subject to} \quad c_1(\underline{x}) \geq 0$$

Let $\underline{x}$ be feasible, and go out to $\underline{x} + \underline{\delta}$ also feasible.

$$c_1(\underline{x}) \geq 0$$

$$c_1(\underline{x} + \underline{\delta}) \geq 0$$

By Taylor expansion:

$$c_1(\underline{x}) + \underline{\delta} \cdot \nabla c_1(\underline{x}) + \frac{1}{2} \delta_i \delta_j \frac{\partial^2 c_1}{\partial x_i \partial x_j} (\underline{x} + \underline{\eta}) \geq 0$$

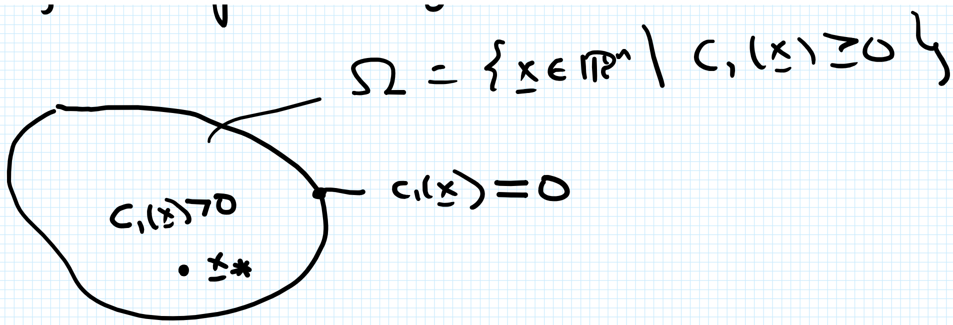
where $\|\underline{\eta}\|_2 < \|\underline{\delta}\|_2$.

For $\|\underline{\delta}\|_2$ sufficiently small, the quadratic term does not contribute to the inequality and we have

$$c_1(\underline{x}) + \underline{\delta} \cdot \nabla c_1(\underline{x}) \geq 0 .$$

Two cases: The minimizer $\underline{x}_*$ is in the interior of the feasible region.

$$\Omega = \{ \underline{x} \in \mathbb{R}^n \mid c_1(\underline{x}) \geq 0 \}$$



$$\underbrace{c_1(\underline{x}_*)}_{>0} + \underline{\delta} \cdot \nabla c_1(\underline{x}_*) \geq 0$$

Thus, the inequality is satisfied for any direction $\underline{\delta}$, once $\|\underline{\delta}\|_2$ is sufficiently small.

Since $\underline{x}_*$ is a minimizer:

$$f(\underline{x}_* + \underline{\delta}) \geq f(\underline{x}_*)$$

$$\Rightarrow \cancel{f(\underline{x}_*)} + \underline{\delta} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} \delta_i \delta_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq \cancel{f(\underline{x}_*)}$$

By TRT, for $\|\underline{y}\|_2 < \|\underline{\delta}\|_2$

$$\Rightarrow \quad \underline{\delta} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} \delta_i \delta_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq 0$$

Choose $\underline{\delta} = t \hat{\delta}$, $\|\hat{\delta}\|_2 = 1$, $t > 0$

$$\Rightarrow t \, \hat{\underline{\delta}} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} t^2 \hat{\delta}_i \hat{\delta}_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq 0$$

Since $t > 0$, divide by t:

$$\hat{\delta} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} t \hat{\delta}_i \hat{\delta}_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq 0$$

$\partial x_i \partial x_j$

Take $t \to 0$:

$$\hat{\delta} \cdot \nabla f(\underline{x}_*) \geq 0 .$$

Since $\underline{x}_*$ is in the interior of the feasible region, there are no constraints on the direction $\hat{\underline{\delta}}$, so choose $\hat{\underline{\delta}} \propto -\nabla f(\underline{x}_*)$

Since $\hat{\delta} \cdot \nabla f(\underline{x}_*) \geq 0$

$$\Rightarrow \; -\nabla f(\underline{x}_*) \cdot \nabla f(\underline{x}_*) \geq 0$$

$$\Rightarrow \; -\|\nabla f(\underline{x}_*)\|_2^2 \geq 0$$

$$\Rightarrow \; \boxed{\nabla f(\underline{x}_*) = 0}$$ CASE 1, $c_1(\underline{x}_*) > 0$

Case 2: The minimizer is on the boundary of the feasible region, $c_1(\underline{x}_*) = 0$.

$$\cancel{\underbrace{c_1(\underline{x}_*)}_{\text{ACTIVE CONSTRAINT}}} + t\hat{\underline{\delta}} \cdot \nabla c_1(\underline{x}_*) \geq 0$$ $\boxed{\underline{\delta} = t\hat{\delta}}$

$$\Rightarrow \; \hat{\delta} \cdot \nabla c_1(\underline{x}_*) \geq 0$$

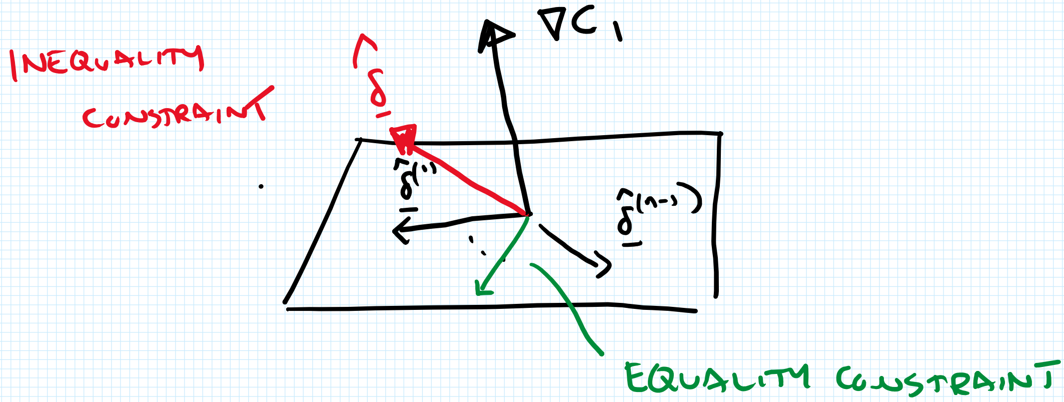
Puts a constraint on the $\hat{\underline{\delta}}$:

$$\hat{\underline{\delta}} \in \mathcal{H}, \quad \mathcal{H} = \left\{ \hat{\underline{\delta}} \in \mathbb{R}^n \;\middle|\; \hat{\delta} \cdot \nabla c_1(\underline{x}_*) \geq 0 \right\}$$

Any $\hat{\underline{\delta}}$ in $\mathcal{H}$ can be decomposed as

$$\hat{\underline{\delta}} = \sum_{i=1}^{n-1} \alpha_i \, \hat{\underline{\delta}}^{(i)} + \beta \, \underline{\nabla} c_1(\underline{x}_*)$$

where $\alpha_i \in \mathbb{R}$ and $\beta > 0$



Since $\underline{x}_*$ is the minimizer:

$$f(\underline{x}_* + \underline{\delta}) \geq f(\underline{x}_*)$$

$$\Rightarrow \quad f(\underline{x}_* + t\hat{\underline{\delta}}) \geq f(\underline{x}_*)$$

TRT

$$\Rightarrow \quad \cancel{f(\underline{x}_*)} + t\hat{\delta} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} t^2 \hat{\delta}_i \hat{\delta}_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq \cancel{f(\underline{x}_*)}$$

where $\|\underline{y}\|_2 \leq \|\underline{\delta}\|_2$, $\quad \underline{\delta} = t\hat{\delta}$, $\quad t > 0$

Divide by $t > 0$:

$$\hat{\delta} \cdot \nabla f(\underline{x}_*) + \frac{1}{2} t \hat{\delta}_i \hat{\delta}_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq 0$$

$$\hat{\delta}\cdot\nabla f(\underline{x}_*) + \frac{1}{2} t \delta_i \delta_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}_* + \underline{y}) \geq 0$$

Take $t \to 0$:

$$\hat{\delta}\cdot\nabla f(\underline{x}_*) \geq 0$$

True for any $\hat{\underline{\delta}} \in \mathcal{H}$.

$$\alpha_i \hat{\underline{\delta}}^{(i)} \cdot \nabla f(\underline{x}_*) \geq 0, \qquad \alpha_i \in \mathbb{R}$$

$$\beta \nabla c_1(\underline{x}_*) \cdot \nabla f(\underline{x}_*) \geq 0, \qquad \beta \geq 0$$

Since the α_i's are pos or neg,

$$\hat{\underline{\delta}}^{(i)} \cdot \nabla f(\underline{x}_*) = 0$$

Only one direction left for ∇f to point in :

$\nabla f \parallel \nabla c_1(\underline{x}_*)$.

Since $\beta \nabla c_1(\underline{x}_*) \cdot \nabla f(\underline{x}_*) \geq 0$, $\beta \geq 0$

$\nabla f(\underline{x}_*)$ must be parallel (not antiparallel) to $\nabla c_1(\underline{x}_*)$.

So, there exists a positive λ_1^* such that

$$\nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*) .$$

Summarizing:

Case 1 : $\left\{ \nabla f(\underline{x}_*) = 0 \right.$

$$\nabla f(\underline{x}_*) = 0$$

<u>Case 1</u> :
$$\begin{cases} \nabla f(\underline{x}_*) = 0 \\ c_1(\underline{x}_*) > 0 \end{cases}$$

INACTIVE CONSTRAINT

<u>Case 2</u> :
$$\begin{cases} \nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*) \\ c_1(\underline{x}_*) = 0 \\ \lambda_1^* \geq 0 \end{cases}$$

ACTIVE CONSTRAINT

Introduce

$$\mathcal{L}(\underline{x}, \lambda) = f(\underline{x}) - \lambda c_1(\underline{x})$$

Bundle both cases together: At the minimizer $\underline{x}_*$, there exists a scalar λ_1^* such that:

KKT CONDITIONS

$$\begin{cases} \nabla f(\underline{x}_*) = \lambda_1^* \nabla c_1(\underline{x}_*) \\ c_1(\underline{x}_*) \geq 0 \\ \lambda_1^* \geq 0 \\ \lambda_1^* c_1(\underline{x}_*) = 0 \end{cases}$$

$$\begin{cases} c_1(\underline{x}_*) = 0 & \text{CASE } 2, \ \lambda_1^* \geq 0 \\ c_1(\underline{x}_*) > 0 & \text{CASE } 1, \ \lambda_1^* = 0 \end{cases}$$

This last one is called the complementarity condition

Today: We have proved the KKT conditions for a

single inequality constraint. ...

Single inequality constraint. Aim of rest of course: extend to a mixture of equality and inequality constraints.