Chapter 3 Numerical Methods

Part 1

3.1 Linearization and Optimization of Functions of Vectors

Problem Notation

Box 3.1: Notation for Numerical Methods Problems

The following notational conventions will be used consistently throughout the text in order to elucidate how most problems reduce to a need for a few fundamental algorithms:

$\underline{x}^* = argmin\,[f(\underline{x})]$	optimization problem
$optimize:_{\underline{x}} \quad f(\underline{x})$	optimization problem
$\underline{g}(\underline{x}) = \underline{b}$	level curve of $\underline{g}(\ )$
$\underline{c}(\underline{x}) = \underline{0}$	rootfinding, constraints
$\underline{z} = \underline{h}(\underline{x})$	measurement of state
$\underline{r}(\underline{x}) = \underline{z} - \underline{h}(\underline{x})$	residual

Outline

- 3.1.1 Linearization
- 3.1.2 Optimization of Objective Functions
- 3.1.3 Constrained Optimization
- Summary

Motivation

- A small number of numerical methods occur frequently.
  - Roots of nonlinear equations
  - Optimization
  - Integration of Diff Eqs.
- Can’t use MATLAB to control the robot.
- You need to know
  - How to implement these.
  - How to cast a problem in standard form.

Motivation

- Techniques will be used for control, perception, position estimation, and mapping.
- Specifically:
  - compute wheel velocities
  - invert dynamic models
  - generate trajectories
  - track features in an image
  - construct globally consistent maps
  - identify dynamic models
  - calibrate cameras

Outline

- 3.1.1 Linearization
- 3.1.2 Optimization of Objective Functions
- 3.1.3 Constrained Optimization
- Summary

3.1.1.1 Taylor Series (About Any Point)

- Scalar function of scalar:

$$f(x+\Delta x) = f(x) + \Delta x \overset{\text{1d}}{\left\{\frac{df}{dx}\right\}_x} + \frac{\Delta x^2}{2!}\overset{\text{1d}}{\left\{\frac{d^2 f}{dx^2}\right\}_x} + \frac{\Delta x^3}{3!}\overset{\text{1d}}{\left\{\frac{d^3 f}{dx^3}\right\}_x} + \ldots$$

- Scalar function of vector:

$$f(\underline{x}+\Delta \underline{x}) = f(\underline{x}) + \Delta \underline{x} \overset{\text{1d}}{\left\{\frac{\partial f}{\partial \underline{x}}\right\}_{\underline{x}}} + \frac{\Delta \underline{x}^2}{2!}\overset{\text{2d}}{\left\{\frac{\partial^2 f}{\partial \underline{x}^2}\right\}_{\underline{x}}} + \frac{\Delta \underline{x}^3}{3!}\overset{\text{3d}}{\left\{\frac{\partial^3 f}{\partial \underline{x}^3}\right\}_{\underline{x}}} + \ldots$$

- Vector function of vector:

$$\underline{f}(\underline{x}+\Delta \underline{x}) = \underline{f}(\underline{x}) + \Delta \underline{x} \overset{\text{2d}}{\left\{\frac{\partial \underline{f}}{\partial \underline{x}}\right\}_{\underline{x}}} + \frac{\Delta \underline{x}^2}{2!}\overset{\text{3d}}{\left\{\frac{\partial^2 \underline{f}}{\partial \underline{x}^2}\right\}_{\underline{x}}} + \frac{\Delta \underline{x}^3}{3!}\overset{\text{4d}}{\left\{\frac{\partial^3 \underline{f}}{\partial \underline{x}^3}\right\}_{\underline{x}}} + \ldots$$

3.1.1.2 Taylor Remainder Theorem

$$f(x + \Delta x) = T_n(\Delta x) + R_n(\Delta x)$$

$$R_n(\Delta x) = \frac{1}{n!}\int_{x}^{(x+\Delta x)} (x - \zeta)^n \left\{ \frac{\partial^{(n)} \underline{f}}{\partial \underline{x}^n} \right\}_{\zeta} d\zeta$$

- Provided the nth derivative is bounded, the error in a truncated Taylor series is of order $(\Delta x)^{n+1}$.
- When Δx is small, $(\Delta x)^{n+1}$ is very small.

3.1.1.3 Linearization

$$\underline{f}(\underline{x} + \Delta\underline{x}) = \underline{f}(\underline{x}) + \Delta\underline{x}\left\{\frac{\partial \underline{f}}{\partial \underline{x}}\right\}_{\underline{x}}$$

- This is accurate “to first order” …
- … meaning the error is second order.

3.1.1.4 Gradients and Level Curves and Surfaces

$$g(\underline{x}) = c \qquad \underline{x} \in \Re^n$$

- This forms an n-1 dimensional subspace of $\mathcal{R}^n$ called a *level surface* (a set of level curves).
- Let the set of solutions be parameterized locally in 1D by the scalar s to form the level curve $\underline{x}(s)$.
  - Each element of $\underline{x}(s)$ is a level curve for which $g(\underline{x})$ is constant. You could plot $x_1(s)$ vs $x_2(s)$for example.
- By the chain rule applied along the constraint:

$$\frac{\partial g(\underline{x}(\underline{s}))}{\partial \underline{s}} = \frac{\partial g}{\partial \underline{x}} \frac{\partial \underline{x}}{\partial \underline{s}}^T = \frac{\partial c}{\partial s} = \underline{0}^T$$

Gradient is normal to all the level curves

3.1.1.5 Example

- Consider the constraint:

$$g(\underline{x}) = x_1^2 + x_2^2 = R^2$$

- The constraint **gradient** is:

$$\frac{\partial g}{\partial \underline{x}} = 2\begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

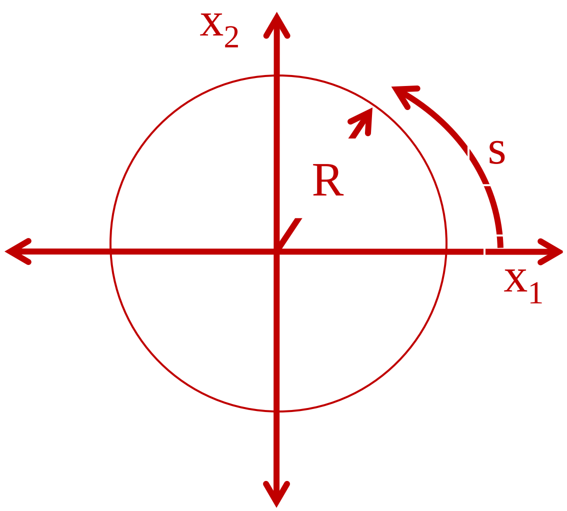
- Parameterize the level curve with:

$$x(s) = \begin{bmatrix} \cos(\kappa s) & \sin(\kappa s) \end{bmatrix}^T$$

- The level curve **tangent**:

$$\frac{\partial \underline{x}}{\partial \underline{s}} = \begin{bmatrix} -\kappa \sin(\kappa s) & \kappa \cos(\kappa s) \end{bmatrix} = \kappa\begin{bmatrix} -x_2 & x_1 \end{bmatrix}$$

- Which is orthogonal to $\partial g / \partial \underline{x}$



3.1.1.6 Jacobians and Level Surfaces

$$\underline{g}(\underline{x}) = \underline{c} \qquad \underline{x} \in \Re^n \qquad \underline{g} \in \Re^m$$

- This forms an n-m dimensional subspace of $\mathcal{R}^n$ called a *level surface* (a set of level curves).
- Let the set of solutions be parameterized by the vector $\underline{s}$ to form $\underline{x}(\underline{s})$.
  - $\underline{x}(s_1)$ and $\underline{x}(s_2)$ are different level curves.
- By the chain rule:

$$\frac{\partial \underline{g}(\underline{x}(\underline{s}))}{\partial \underline{s}} = \frac{\partial \underline{g}}{\partial \underline{x}} \frac{\partial \underline{x}}{\partial \underline{s}}^T = \frac{\partial \underline{c}}{\partial \underline{s}} = [0]$$

Rows of Jacobian are orthogonal to tangents to level curves

Jacobian and Tangent Plane

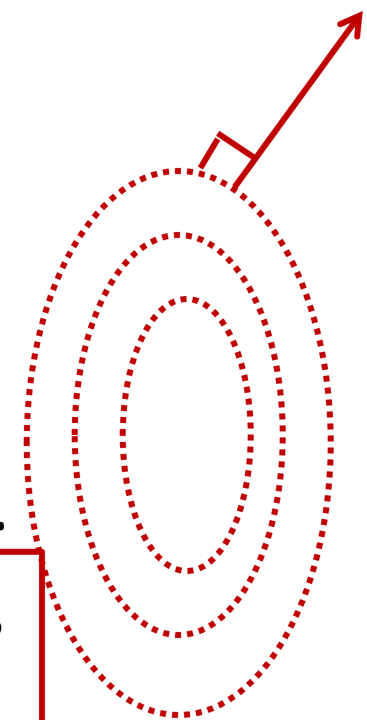
- Consider again:

Surface Tangents

$$\frac{\partial \underline{g}}{\partial \underline{x}} \frac{\partial \underline{x}}{\partial \underline{s}}^{T} = [0]$$

Jacobian

- The set of all linear combinations of the rows of $\partial \underline{x} / \partial \underline{s}$ is called the *tangent plane*.
- Therefore, the constraint tangent plane lies in the nullspace of the constraint Jacobian.



Outline

- 3.1.1 Linearization
- <u>3.1.2 Optimization of Objective Functions</u>
- 3.1.3 Constrained Optimization
- Summary

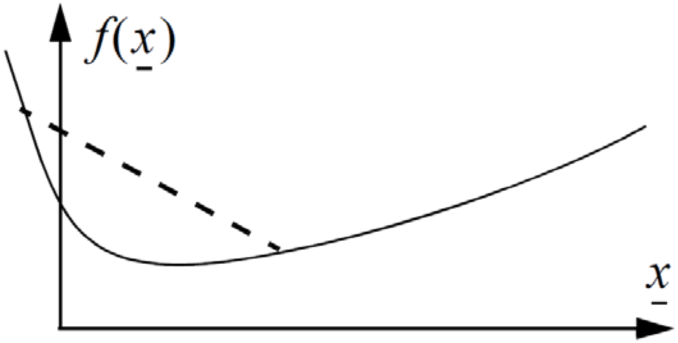
3.1.2 Optimization of Objective Functions

$$\text{optimize:}_{\underline{x}} \quad f(\underline{x}) \qquad\qquad \underline{x} \in \Re^n$$

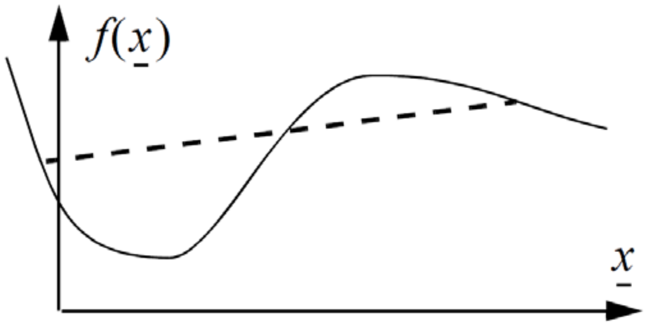
- f is always a scalar.
- Called *objective, cost* or *utility* function
- The max(f) problem is equivalent to min(-f) so we will treat all problems as minimization.

Convexity and Global Minima

- A convex (concave) function is uniformly on or below (above) any line between any two points in its domain.
- For these, <u>local and global minima are the same</u>.

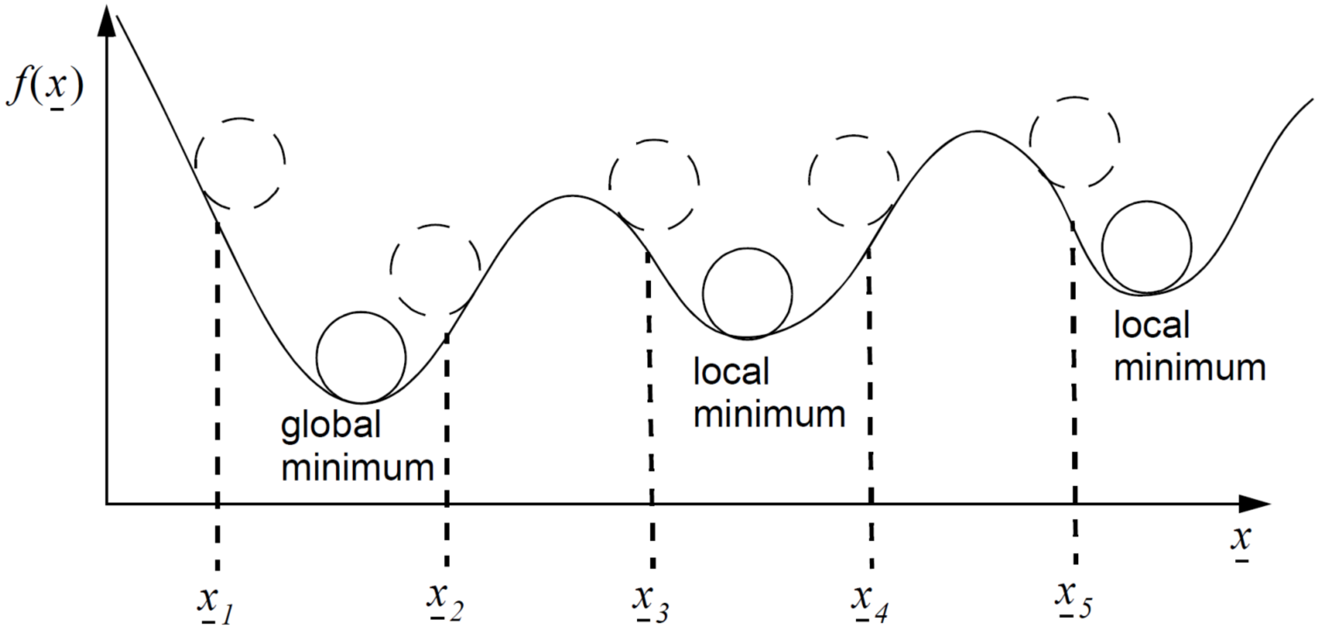


Convex



Not Convex

3.1.2.1 Local Minima



- Suppose we compute a minimum over a specified region
- Which one you get often depends on where you start searching.
- Any global minimum (not on a boundary of the set) must also be a local minimum.
- Therefore a local minimum is a more fundamental problem.

3.1.2.1 Notation

- This section will henceforth use two forms of shorthand in one compact form:

$$\underline{f}_{\underline{x}}(\underline{x}^*) \equiv \frac{\partial \underline{f}(\underline{x}^*)}{\partial \underline{x}} \equiv \left. \frac{\partial \underline{f}(\underline{x})}{\partial \underline{x}} \right|_{\underline{x} = \underline{x}^*}$$

- So $\underline{f}_{\underline{x}}()$ is a Jacobian or a gradient (if f is scalar-valued):

3.1.2.1 First Order (Necessary) Conditions

- Consider satisfying the weaker condition of being a local minimum.
- In a given neighborhood, a Taylor series approximation applies.
- Consider the change Δf after moving a distance $\Delta \underline{x}$ from a local minimum $\underline{x}^*$.

$$\Delta f = f(\underline{x}^* + \Delta\underline{x}) - f(\underline{x}^*) = f(\underline{x}^*) + \left\{\frac{\partial f(\underline{x}^*)}{\partial \underline{x}}\right\}\Delta\underline{x} - f(\underline{x}^*)$$

$$\Delta f = \left\{\frac{\partial f(\underline{x}^*)}{\partial \underline{x}}\right\}\Delta\underline{x} = f_{\underline{x}}(\underline{x}^*)\Delta\underline{x}$$

Objective Gradient

3.1.2.1 First Order (Necessary) Conditions

- From last slide… $\Delta f = f_{\underline{x}}(\underline{x}^*)\Delta\underline{x}$
- If $\underline{x}^*$ is a local minimum then $\Delta f \geq 0$ for an arbitrary $\Delta\underline{x}$.
- But if $\Delta f(\Delta\underline{x}) > 0$ then $\Delta f(-\Delta\underline{x})$ would be < 0. Contradiction.
- Hence we must have equality …

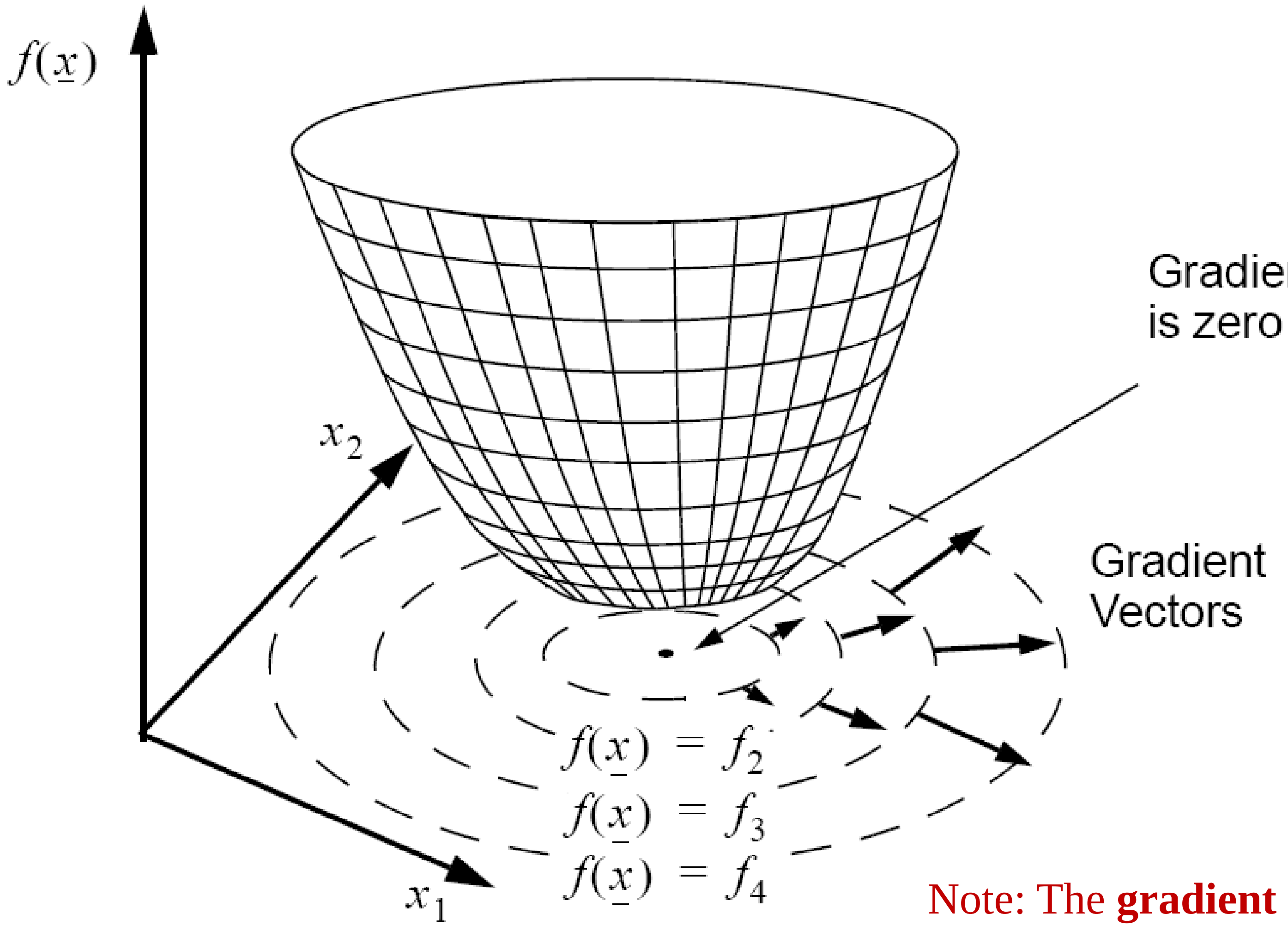
$$\Delta f = f_{\underline{x}}(\underline{x}^*)\Delta\underline{x} = 0 \qquad \forall \Delta\underline{x}$$

- But since Δx is arbitrary, we must have:

$$\boxed{f_{\underline{x}}(\underline{x}^*) = \underline{0}^T}$$

- … at a local minimum.

3.1.2.1 First Order (Necessary) Conditions



Note: The **gradient exists in the x1-x2 plane**, not in the x1-x2-f volume.

3.1.2.2 Second Order (Sufficient) Conditions

- At a local minimum, a perturbation in any direction satisfies:

$$f(\underline{x}^* + \Delta\underline{x}) = f(\underline{x}^*) + f_{\underline{x}}(\underline{x}^*)\Delta\underline{x} = f(\underline{x}^*) \qquad \forall \Delta\underline{x}$$

- Therefore, the order 2 Taylor series has no linear term at this point:
  - Instead, it looks like:

$$f(\underline{x} + \Delta\underline{x}) = f(\underline{x}) + \frac{\Delta\underline{x}^2}{2!}\left\{\frac{\partial^2 f}{\partial \underline{x}^2}\right\}_{\underline{x}}$$

Implicit Layout

3.1.2.2 Second Order (Sufficient) Conditions

- In explicit vector layout this is:

$$\Delta f = f(\underline{x} + \Delta\underline{x}) - f(\underline{x}) = f(\underline{x}) + \frac{1}{2}\Delta\underline{x}^T F_{\underline{x}\underline{x}}(\underline{x})\Delta\underline{x}$$

- So the optimality condition is:

$$\Delta f = \frac{1}{2}\Delta\underline{x}^T F_{\underline{x}\underline{x}}(\underline{x})\Delta\underline{x} \geq 0 \qquad \forall \Delta\underline{x}$$

- In other words, the Hessian must be positive semi-definite.

- The sufficient condition is:

$$\boxed{\Delta\underline{x}^T F_{\underline{x}\underline{x}}(\underline{x})\Delta\underline{x} > 0}$$

Hessian Must Be Positive Definite

Feasible Points

- Consider a set of m nonlinear constraints on the n elements of $\underline{x}$.

$$\underline{c}(\underline{x}) = \underline{0} \quad \underline{c} \in \Re^{m} \quad \underline{x} \in \Re^{n}$$

- Any point that satisfies these is called a <u>feasible</u> point.
- Suppose the point $\underline{x}'$ is feasible.

3.1.2.3 Feasible Perturbations

- Consider a feasible perturbation Δx
  - That is, one which remains feasible to first order.

$$\underline{c}(\underline{x}' + \Delta\underline{x}) = \underline{c}(\underline{x}') + \underline{c}_{\underline{x}}(\underline{x}')\Delta\underline{x} = \underline{0}$$

- The Jacobian $\underline{c}_{\underline{x}}$ is nonsquare m by n.
- Since $\underline{x}'$ is feasible, $\underline{c}(\underline{x}')=0$. Hence:

$$\boxed{\underline{c}_{\underline{x}}(\underline{x}')\Delta\underline{x} = \underline{0}}$$

- For a feasible perturbation from a feasible point.

3.1.2.4 Constraint Nullspace

- From last slide $\underline{c}_{\underline{x}}(\underline{x}')\Delta\underline{x} = \underline{0}$
- The Jacobian is composed of m gradients (rows):

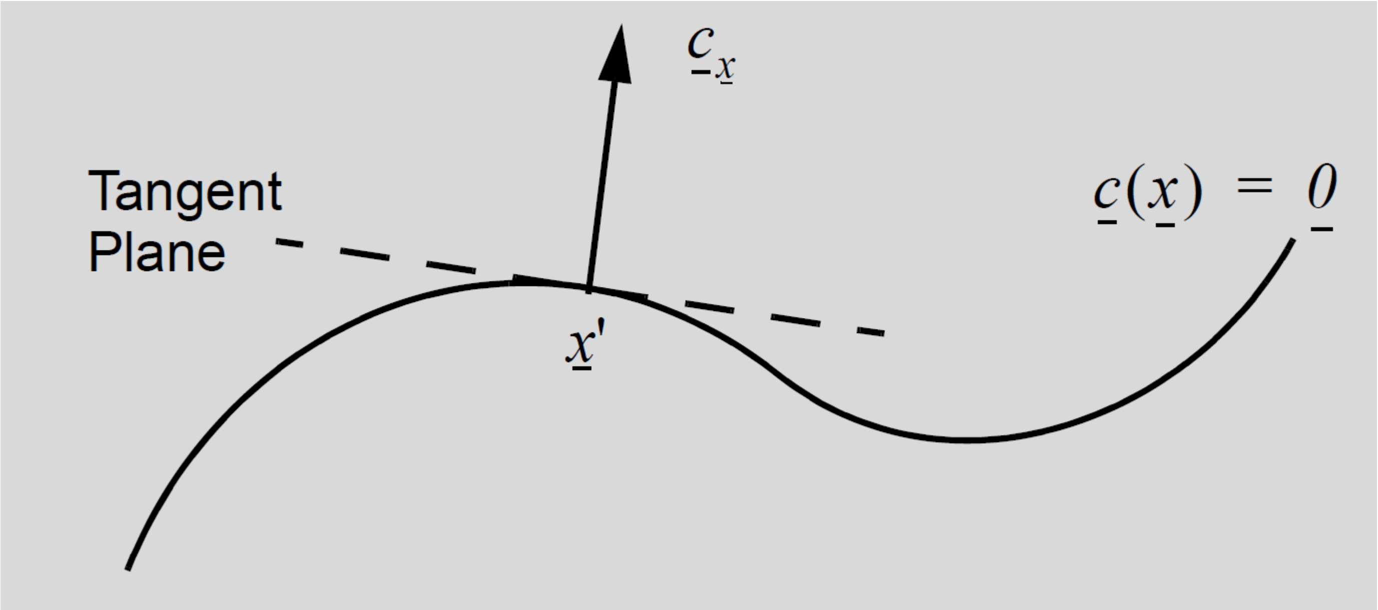
$$\{ c_1(\underline{x}), c_2(\underline{x}) \ldots c_m(\underline{x}) \}$$

- The feasibility condition is requiring the perturbation to be in the *nullspace* of the constraints.

$$\boxed{\mathcal{N}(\underline{c}_{\underline{x}}) = \{ \Delta\underline{x} : \quad \underline{c}_{\underline{x}}(\underline{x}')\Delta\underline{x} = \underline{0} \}}$$

3.1.2.4 Constraint Tangent Plane

- The constraint nullspace is also known as the constraint tangent plane.



Conditions for Optimality and Feasibility

Box 3.2: Conditions for Optimality and Feasibility

For the scalar-valued objective function $f(\underline{x})$ ...

The necessary condition for a local optimum at $\underline{x}^*$ is that the gradient with respect to $\underline{x}$ vanish at that point:

$$f_{\underline{x}}(\underline{x}^*) = \underline{0}^T$$

The sufficient condition for a local minimum at $\underline{x}^*$ is that the Hessian be positive definite:

$$\frac{1}{2}\Delta\underline{x}^T f_{\underline{xx}}(\underline{x})\Delta\underline{x} > 0 \qquad \forall \Delta\underline{x}$$

For the set of constraint functions $\underline{c}(\underline{x}) = \underline{0}$, if $\underline{x}'$ is a feasible point (i.e., $\underline{c}(\underline{x}') = \underline{0}$, then the new point $\underline{x}'' = \underline{x}' + \Delta\underline{x}$ is also feasible (to first order) if the state change $\Delta\underline{x}$ is confined to the constraint tangent plane.

$$\underline{c}_{\underline{x}}(\underline{x}')\Delta\underline{x} = \underline{0}$$

Outline

- 3.1.1 Linearization
- 3.1.2 Optimization of Objective Functions
- <u>3.1.3 Constrained Optimization</u>
- Summary

3.1.3 Constrained Optimization

$$\text{optimize:}_{\underline{x}} \quad f(\underline{x}) \qquad\qquad \underline{x} \in \Re^n$$

$$\text{subject to:} \quad \underline{c}(\underline{x}) = \underline{0} \qquad\qquad \underline{c} \in \Re^m$$

- n variables in f() subject to m<n constraints.
- Also known as *nonlinear programming*.
- If there is no f(), it is *constraint satisfaction* or *rootfinding*.
- If there are no constraints it is *unconstrained optimization*.

3.1.3.1 First Order (Necessary) Conditions

- There are n-m degrees of freedom left after the constraints are enforced.
- Cannot just set objective gradient to zero because such a point may not be feasible.
- Define a local minimum to mean …
  - … for a feasible perturbation from a feasible point…
  - … the objective must not increase …

$$\boxed{\begin{array}{ll} \underline{c}_{\underline{x}}(\underline{x}^{*})\Delta\underline{x} = \underline{0} & \forall\Delta\underline{x}\ \text{feasible} \\ f_{\underline{x}}(\underline{x}^{*})\Delta\underline{x} = 0 & \forall\Delta\underline{x}\ \text{feasible} \end{array}}$$

Lagrange Multipliers

- First, a feasible perturbation lies in the constraint tangent plane.

$$\underline{c}_{\underline{x}}(\underline{x}^*)\Delta\underline{x} = \underline{0} \qquad \forall \Delta\underline{x} \text{ feasible}$$

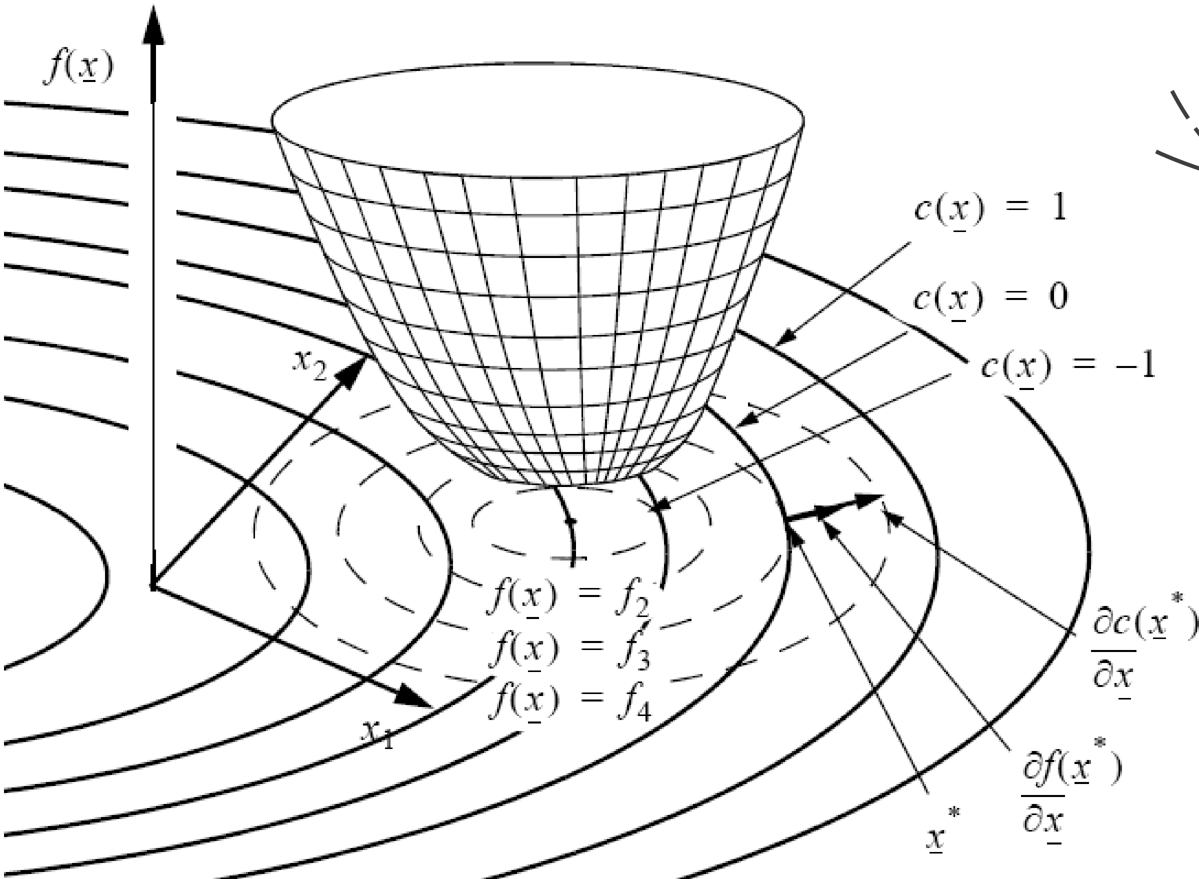
- Second, the objective gradient is orthogonal to an arbitrary vector in the tangent plane.

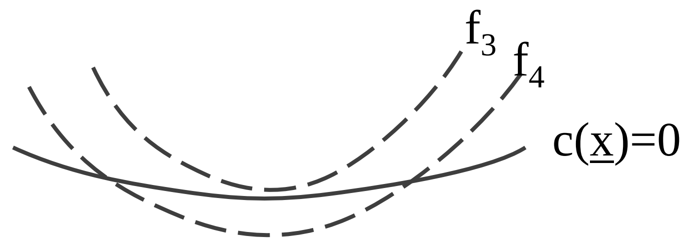
$$f_{\underline{x}}(\underline{x}^*)\Delta\underline{x} = 0 \qquad \forall \Delta\underline{x} \text{ feasible}$$

- The rows of $\underline{c}_{\underline{x}}(\underline{x}^*)$ span the space of all vectors orthogonal to the constraints.
- Hence, $f_{\underline{x}}()$ must be some linear combination of them:

$$\boxed{f_{\underline{x}}(\underline{x}^*) = \underline{\lambda}^T \underline{c}_{\underline{x}}(\underline{x}^*)}$$

Constrained Local Minimum





Moving left or right on the constraint increases the objective

- The objective gradient is orthogonal to the level curves of both the constraints and the objective.
- Level curves of constraints and objective are mutually tangent.

Necessary Conditions

- Reverse the sign of the (unknown) multipliers to get:

$$\underline{f}_{\underline{x}}^{T}(\underline{x}^{*}) + \underline{c}_{\underline{x}}^{T}(\underline{x}^{*})\underline{\lambda} = \underline{0}$$

$$\underline{c}(\underline{x}) = \underline{0}$$

Eqn A

- It is customary to define the *Lagrangian*:

$$l(\underline{x}, \underline{\lambda}) = f(\underline{x}) + \underline{\lambda}^{T}\underline{c}(\underline{x})$$

Necessary Conditions

- Define the *Lagrangian*:

$$l(\underline{x}, \underline{\lambda}) = f(\underline{x}) + \underline{\lambda}^T \underline{c}(\underline{x})$$

- The necessary conditions (Eqn A) can then be written as:

$$\boxed{\begin{aligned} \frac{\partial l(\underline{x}, \underline{\lambda})}{\partial \underline{x}}^T &= \underline{0} \\ \frac{\partial l(\underline{x}, \underline{\lambda})}{\partial \underline{\lambda}}^T &= \underline{0} \end{aligned}}$$

Necessary Conditions For A Constrained Local Mimimum

3.1.3.2 Solving NLP Problems

- Four basic techniques:
  - Substitution: Substitute constraints into objective. Solve resulting n-m dof unconstrained problem.
  - Descent: Follow negative gradient of f() while remaining feasible.
  - Lagrange Multipliers: Solve n+m (linearized) necessary conditions directly.
  - Penalty Function: Convert constraints to a cost and solve resulting n+m dof unconstrained problem.

3.1.3.5 Example: Penalty Function

- Consider the problem:

$$\text{minimize:}_{\underline{x}} \quad f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} \qquad \underline{x} \in \Re^n$$
$$\text{subject to:} \quad G\underline{x} = \underline{b} \qquad \underline{b} \in \Re^m$$

- Suppose constraints are not satisfied and define the residual:

$$\underline{r}(\underline{x}) = G\underline{x} - \underline{b} \qquad \underline{r} \in \Re^m$$

- Define a weight matrix R and form the new cost function:

$$\text{minimize:}_{\underline{x}} \quad f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} + \frac{1}{2}\underline{r}^T(\underline{x}) R \underline{r}(\underline{x}) \qquad \underline{x} \in \Re^n$$

Example: Penalty Function

- From last slide:

$$\text{minimize:}_{\underline{x}} \quad f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} + \frac{1}{2}\underline{r}^T(\underline{x}) R \underline{r}(\underline{x}) \qquad \underline{x} \in \Re^n$$

- Differentiate f(x) wrt x and solve:

$$f_{\underline{x}} = \underline{x}^T Q + \underline{r}^T(\underline{x}) R \frac{\partial \underline{r}(\underline{x})}{\partial \underline{x}} = \underline{0}^T$$

$$f_{\underline{x}} = \underline{x}^T Q + (\underline{x}^T G^T - \underline{b}^T) R G = \underline{0}^T$$

$$\underline{x}^T Q + \underline{x}^T G^T R G = \underline{b}^T R C$$

$$\{Q + G^T R G\}\underline{x} = G^T R \underline{b}$$

- Hence:

$$\boxed{\underline{x} = \{Q + G^T R C\}^{-1} G^T R \underline{b}}$$

Solution Depends on Weight Matrix

Outline

- 3.1.1 Linearization
- 3.1.2 Optimization of Objective Functions
- 3.1.3 Constrained Optimization
- <u>Summary</u>

Summary

- Linearization, derived from the Taylor series, is the fundamental basis of many numerical methods.
- The gradient vanishes at a local minimum of an unconstrained objective.
- All the gradient(s) of a set of constraints are orthogonal to every vector in the constraint tangent plane.

Summary

- A constrained minimum occurs when the objective tangent plane is tangent to the constraint tangent plane.
  – Equivalently, when the objective gradient is a linear combination of the constraint gradients.
- The Lagrange multipliers are a set of unknown projections of a vector onto a set of constraint gradients (disallowed directions).