Optimal Design : Boundary Optima

ME 7129
Kuei-Yuan Chan
National Taiwan University

When constraints are present, monotonic behaviors often result in optimal solutions being on the boundary of constraints. However, monotonicity analysis is seldom done without iteration. In addition, equality constraints of implicit functions can not be used directly for dimension reduction.

Infeasible Design by Decent Direction

- For unconstrained problems, $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{s}_k$ is assumed to be still an interior point.
- For constrained problems, this assumption can be violated.
- Feasible perturbation :

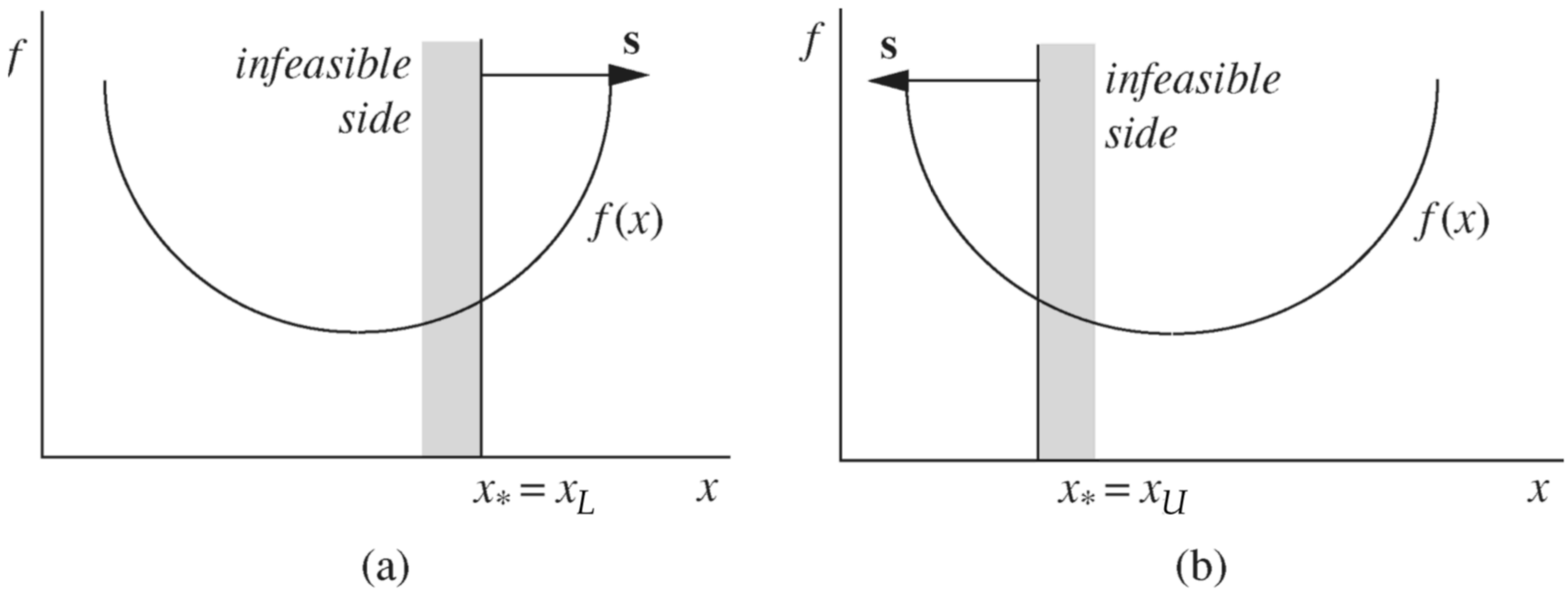
$\partial \mathbf{x}_k = \alpha_k \mathbf{s}_k$ is a feasible perturbation if and only if $\mathbf{x} + \partial \mathbf{x} \in \mathcal{X}$

Necessary Condition

- A local minimum must satisfy the condition that any feasible perturbation will result in objective function increase.
- That is

$$\nabla f(\mathbf{x}_*)\mathbf{s} \geq 0 \quad \text{for all feasible } \mathbf{s}$$

One-Dimension Demonstration

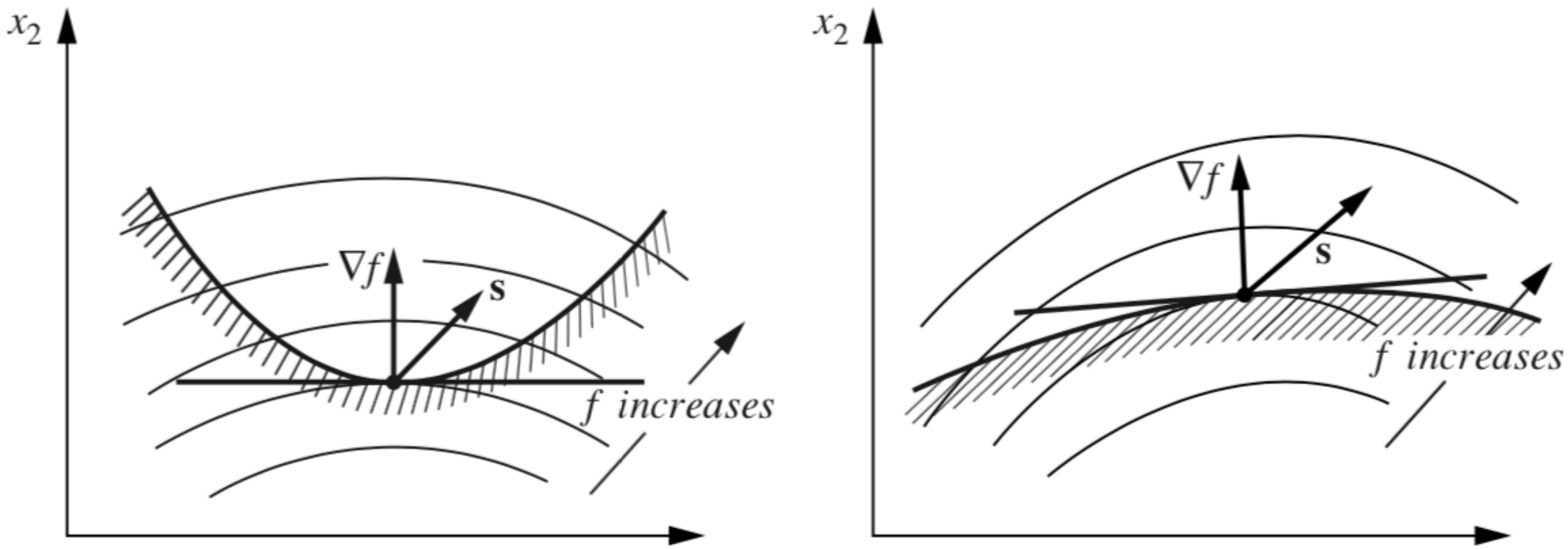


at a local boundary minimum the slope of the function and the allowable change in x must have the same sign

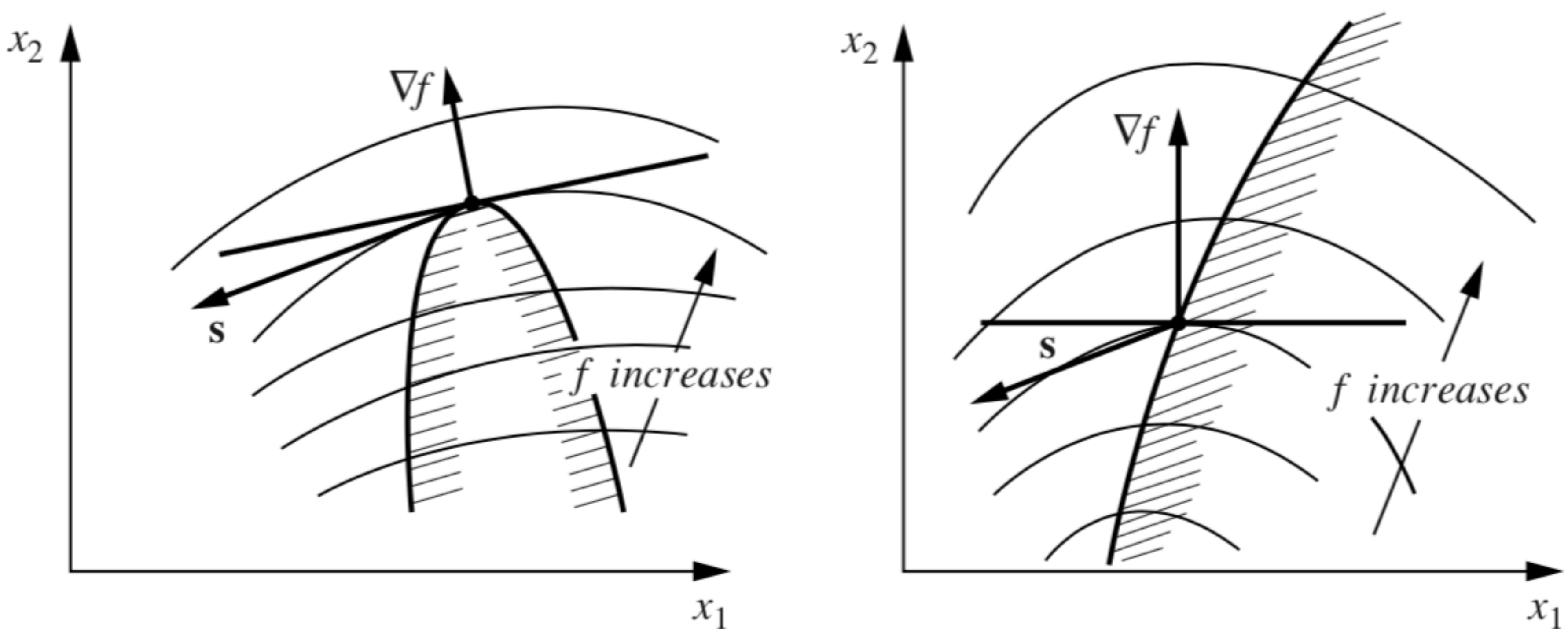
$$\{\min f(x^+),\ \text{subject to } g(x^-) = x_L - x \leq 0\}$$

ME7129
K.-Y. Chan

Two-Dimension Demonstration



at a minimum no feasible directions should exist that are at an angle of more than 90° from the gradient direction



(c) (d)

Condition (5.1) implies a fundamental way for constructing iterative procedures to solve constrained problems. If $\mathbf{x}_k$ is a nonoptimal point, a move in the feasible direction $\mathbf{s}_k$ should be made so that $\nabla f(\mathbf{x}_k)\mathbf{s}_k < 0$. The step length α_k is found from solving the problem

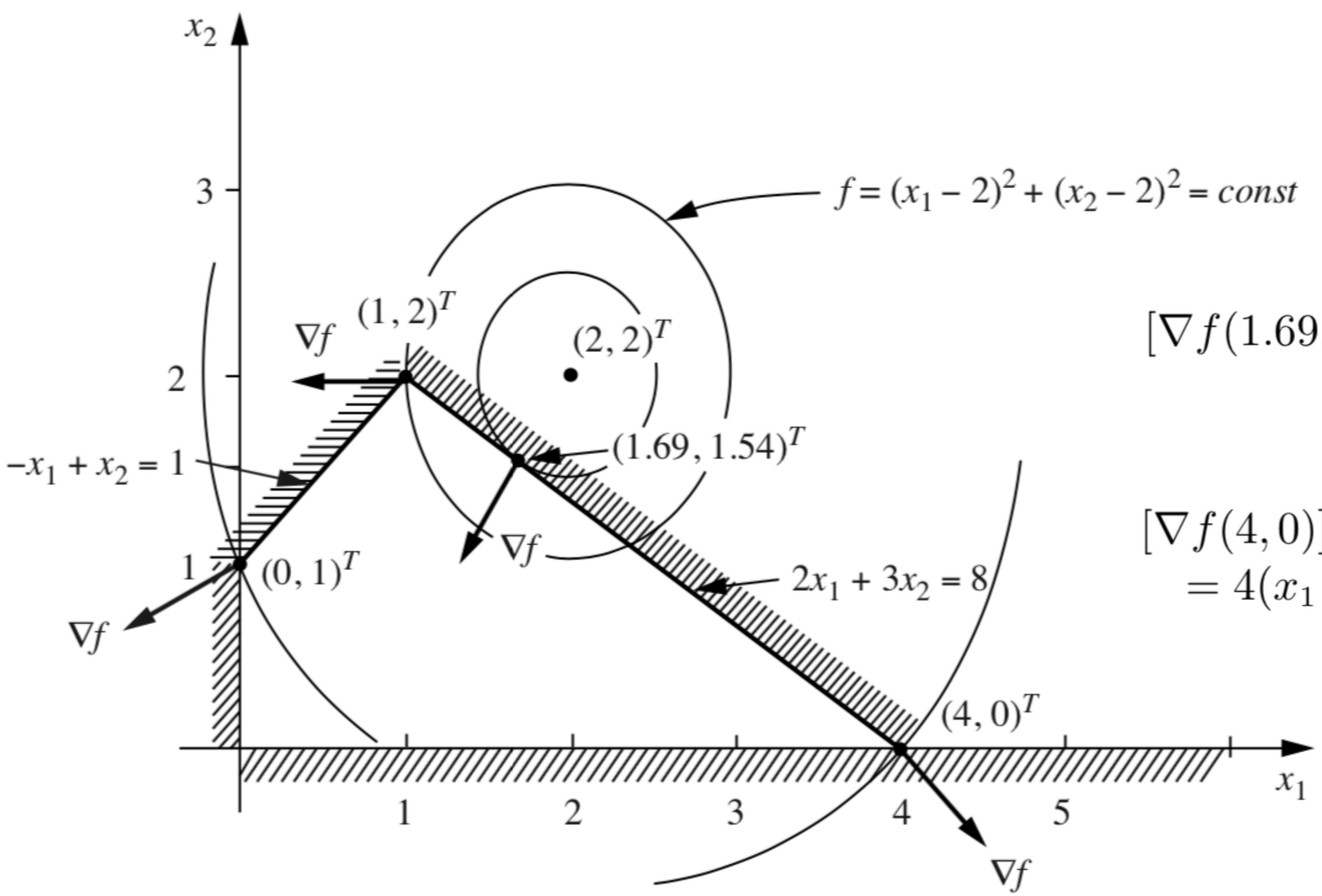
$$\min_{0\leq\alpha<\infty} f(\mathbf{x}_k + \alpha\mathbf{s}_k), \quad \text{subject to } \mathbf{x}_k + \alpha\mathbf{s} \in \mathcal{X}. \tag{5.3}$$

Example 5.1 Consider minimizing the function

$$f = (x_1 - 2)^2 + (x_2 - 2)^2 \quad \text{where } (x_1, x_2)^T \text{ belongs to the set } \mathcal{X} \subseteq R^n$$

defined by

$$\mathcal{X} = \{\mathbf{x} \mid -x_1 + x_2 \le 1, 2x_1 + 3x_2 \le 8,\ x_1 \ge 0,\ x_2 \ge 0\}.$$



$$[\nabla f(1.69, 1.54)][(x_1, x_2)^T - (1.69, 1.54)^T] \geq 0$$

$$[\nabla f(4, 0)][(x_1, x_2)^T - (4, 0)^T] = (4, -4)(x_1 - 4, x_2)^T$$
$$= 4(x_1 - x_2 - 4) \leq 0 \quad \text{for all feasible } x_1, x_2.$$

A set of equality constraints on $\Re^n$, $h_1(\mathbf{x}) = 0$ $, \cdots, h_m(\mathbf{x}) = 0$ defines a hypersurface of dimension $n - m$ if the constraints are functionally independent.

Regularity

- A point is called a regular point of and only if the constraint gradient vectors are linearly independent.
- The assumption that any point we consider must be regular is often referred to as constraint qualification.

Normal Plane

The normal plane for $\mathcal{S}$ at a regular point $\mathbf{x}$ is the subspace $\mathcal{N}(\mathbf{x})$ of $\Re^n$ spanned by the gradient vectors $\nabla h_j(\mathbf{x})$

$$\mathcal{N}(\mathbf{x}) = \{\mathbf{z} \in \Re^n : \alpha_1, \cdots, \alpha_m \in \Re;$$
$$\mathbf{z} = \alpha_1 \nabla h_1^T(\mathbf{x}) + \cdots + \alpha_m \nabla h_m^T(\mathbf{x})$$

Tangent Plane

The tangent plane for $\mathcal{S}$ at a regular point $\mathbf{x}$ is the subspace $\mathcal{T}(\mathbf{x})$ of $\Re^n$, orthogonal to the normal space. That is

$$\mathcal{T}(\mathbf{x}) = \{\mathbf{y} \in \Re^n : \nabla \mathbf{h}(\mathbf{x})\mathbf{y} = \mathbf{0}\}$$

Note

- The condition of regularity is not imposed on the constraint surface itself but on its representation in terms of a constraint set.
- The above definition of normal and tangent planes require that they pass through the origin. It is conceptually better to think that they pass through the design point.
- In the neighborhood of $\mathbf{x}_0$, we can represent points $\mathbf{x}_1$ by moving along the tangent and normal spaces. (see Fig.5.4)

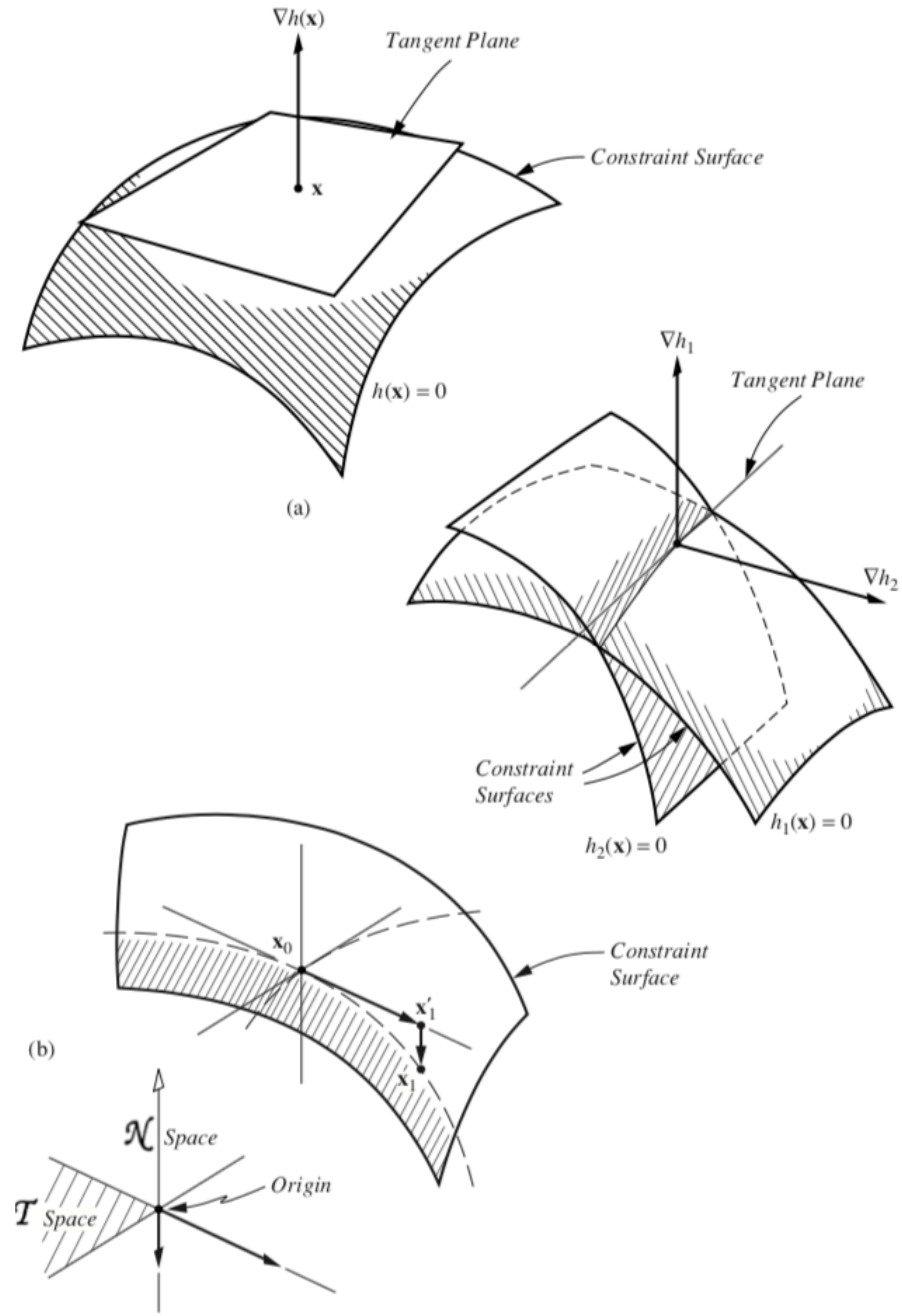


Figure 5.4. (a) Tangent planes; (b) representation of points in the neighborhood of $\mathbf{x}_0$.

Dealing with Equality Constraints

- The simplest way in dealing with equality constraints is via direct elimination.
- Although the method results in reduced problem and can be effective, we still need some other method when constraints are too difficult to solve explicitly.

Reduced Gradient

- Let us now define the necessary condition for constrained optimization problems.
- Consider the equality constrained problem :

$$\min f(\mathbf{x})$$
$$\text{subject to} \quad \mathbf{h}(\mathbf{x}) = 0$$

The first-order approximations of the perturbations for objective and constraint functions are

$$\partial f = \nabla f \partial \mathbf{x} = \sum_{i=1}^{n} (\partial f / \partial x_i) \partial x_i, \tag{5.9}$$
$$\partial h_j = \nabla h_j \partial \mathbf{x} = \sum_{i=1}^{n} (\partial h_j / \partial x_i) \partial x_i = 0, \quad j = 1, 2, \ldots, m.$$

Reduced Gradient (cont.)

We can rearrange (5.9) as follows:

$$-\partial f + \sum_{i=1}^{m} (\partial f / \partial x_i) \partial x_i = - \sum_{i=m+1}^{n} (\partial f / \partial x_i) \partial x_i, \tag{5.10}$$

$$\sum_{i=1}^{m} (\partial h_j / \partial x_i) \partial x_i = - \sum_{i=m+1}^{n} (\partial h_j / \partial x_i) \partial x_i, \quad j = 1, \ldots, m.$$

define state variables

$$s_i \triangleq x_i; \quad i = 1, \ldots, m.$$

define decision variables

$$d_i \triangleq x_i \quad i = m+1, \ldots, n \quad \text{or} \quad i = 1, \ldots, p.$$

$$-\partial f + \sum_{i=1}^{m} (\partial f / \partial s_i) \partial s_i = - \sum_{i=1}^{p} (\partial f / \partial d_i) \partial d_i,$$

$$\sum_{i=1}^{m} (\partial h_j / \partial s_i) \partial s_i = - \sum_{i=1}^{p} (\partial h_j / \partial d_i) \partial d_i, \quad j = 1, \ldots, m.$$

in vector form

$$-\partial f + (\partial f / \partial \mathbf{s}) \partial \mathbf{s} = -(\partial f / \partial \mathbf{d}) \partial \mathbf{d},$$

$$(\partial \mathbf{h} / \partial \mathbf{s}) \partial \mathbf{s} = -(\partial \mathbf{h} / \partial \mathbf{d}) \partial \mathbf{d},$$

Reduced Gradient (cont.)

$$-\partial f + (\partial f/\partial \mathbf{s})\partial \mathbf{s} = -(\partial f/\partial \mathbf{d})\partial \mathbf{d},$$
$$(\partial \mathbf{h}/\partial \mathbf{s})\partial \mathbf{s} = -(\partial \mathbf{h}/\partial \mathbf{d})\partial \mathbf{d},$$

$$\partial \mathbf{s} = -(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d})\partial \mathbf{d},$$
$$\partial f = (\partial f/\partial \mathbf{d})\partial \mathbf{d} + (\partial f/\partial \mathbf{s})\partial \mathbf{s}$$
$$= [(\partial f/\partial \mathbf{d}) - (\partial f/\partial \mathbf{s})(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d})]\partial \mathbf{d}.$$

The quantity in brackets can be thought of as the gradient of a new *unconstrained* function $z(\mathbf{d})$, which would be equivalent to the original objective function f if the solution variables had been eliminated. Thus, we can define a quantity

$$\partial z/\partial \mathbf{d} \triangleq (\partial f/\partial \mathbf{d}) - (\partial f/\partial \mathbf{s})(\partial \mathbf{h}/\partial \mathbf{s})^{-1}(\partial \mathbf{h}/\partial \mathbf{d}), \tag{5.16}$$

Lagrange Multipliers

the necessary condition $(\partial z/\partial \mathbf{d})_\dagger = \mathbf{0}^T,$ becomes

$$\left(\frac{\partial f}{\partial \mathbf{d}}\right)_\dagger - \underline{\left(\frac{\partial f}{\partial \mathbf{s}}\right)_\dagger \left(\frac{\partial \mathbf{h}}{\partial \mathbf{s}}\right)_\dagger^{-1}} \left(\frac{\partial \mathbf{h}}{\partial \mathbf{d}}\right)_\dagger = \mathbf{0}^T,$$

$$\lambda^T \triangleq -\left(\frac{\partial f}{\partial \mathbf{s}}\right)_\dagger \left(\frac{\partial \mathbf{h}}{\partial \mathbf{s}}\right)_\dagger^{-1}.$$

Lagrange Function

$$L(\mathbf{x}, \lambda) \triangleq f(\mathbf{x}) + \lambda^T \mathbf{h}(\mathbf{x}),$$

Necessary Condition

$$\nabla f(\mathbf{x}_\dagger) + \lambda^T \nabla \mathbf{h}(\mathbf{x}_\dagger) = \mathbf{0}^T.$$

Constrained Hessian

- A Hessian can generally be represented as

$$(\partial^2\mathbf{y}/\partial\mathbf{x}^2)\triangleq\left(\frac{\partial^2 y_1}{\partial\mathbf{x}^2},\ldots,\frac{\partial^2 y_m}{\partial\mathbf{x}^2}\right)^T \quad \text{for}\, \mathbf{y}=(y_1,\ldots,y_m)^T.$$

- With the reduced gradient, the reduced Hessian is then

$$\begin{aligned}\frac{\partial^2 z}{\partial\mathbf{d}^2}&=\frac{\partial}{\partial\mathbf{d}}\left(\frac{\partial z}{\partial\mathbf{d}}\right)^T=\frac{\partial}{\partial\mathbf{d}}\left(\frac{\partial f}{\partial\mathbf{d}}\right)^T+\frac{\partial}{\partial\mathbf{d}}\left[\left(\frac{\partial f}{\partial\mathbf{s}}\right)\left(\frac{\partial\mathbf{s}}{\partial\mathbf{d}}\right)\right]^T\\&=\frac{\partial^2 f}{\partial\mathbf{d}^2}+\left(\frac{\partial^2 f}{\partial\mathbf{d}\partial\mathbf{s}}\right)\left(\frac{\partial\mathbf{s}}{\partial\mathbf{d}}\right)+\left(\frac{\partial\mathbf{s}}{\partial\mathbf{d}}\right)^T\left[\frac{\partial}{\partial\mathbf{d}}\left(\frac{\partial f}{\partial\mathbf{s}}\right)^T\right]\\&\quad+\left(\frac{\partial f}{\partial\mathbf{s}}\right)\left[\frac{\partial}{\partial\mathbf{d}}\left(\frac{\partial\mathbf{s}}{\partial\mathbf{d}}\right)^T\right]^T\end{aligned}$$

Second-Order Sufficiency

- A feasible point $\mathbf{x}_* = (\mathbf{d}_*, \mathbf{s}_*)^T$ that satisfies the conditions $(\partial z/\partial \mathbf{d})_* = \mathbf{0}^T$ and that $\partial \mathbf{d}^T (\partial^2 z/\partial \mathbf{d}^2)_* \partial \mathbf{d} > 0$ is a local constrained minimum.

Differential Quadratic Form

- Let us define the following shorthand symbols

$$\mathbf{S}_{\mathbf{d}} - \partial \mathbf{s}/\partial \mathbf{d}, \quad L_{\mathbf{dd}} = \partial^2 L/\partial \mathbf{d}^2, \quad L_{\mathbf{ds}} = \partial^2 L/\partial \mathbf{d}\partial \mathbf{s}$$

- After derivation, we can find that the differential quadratic form of the reduced function is equal to the differential quadratic form of the Lagrangian

$$\begin{aligned}
\partial \mathbf{d}^T(\partial^2 z/\partial \mathbf{d}^2)\partial \mathbf{d} &= \partial \mathbf{d}^T(L_{\mathbf{dd}} + \mathbf{S}_{\mathbf{d}}^T L_{\mathbf{sd}} + L_{\mathbf{ds}}\mathbf{S}_{\mathbf{d}} + \mathbf{S}_{\mathbf{d}}^T L_{\mathbf{ss}}\mathbf{S}_{\mathbf{d}})\partial \mathbf{d} \\
&= \partial \mathbf{d}^T L_{\mathbf{dd}}\partial \mathbf{d} + \partial \mathbf{s}^T L_{\mathbf{sd}}\partial \mathbf{d} + \partial \mathbf{d}^T L_{\mathbf{ds}}\partial \mathbf{s} + \partial \mathbf{s}^T L_{\mathbf{ss}}\partial \mathbf{s} \\
&= (\partial \mathbf{d}^T, \partial \mathbf{s}^T)\begin{pmatrix} L_{\mathbf{dd}} & L_{\mathbf{ds}} \\ L_{\mathbf{sd}} & L_{\mathbf{ss}} \end{pmatrix}\begin{pmatrix} \partial \mathbf{d} \\ \partial \mathbf{s} \end{pmatrix} = \partial \mathbf{x}^T L_{\mathbf{xx}}\partial \mathbf{x},
\end{aligned}$$

If a feasible point $\mathbf{x}_*$ *exists together with a vector* λ *such that* $\nabla f(\mathbf{x}_*) + \lambda^T \nabla \mathbf{h}(\mathbf{x}_*) = \mathbf{0}$ *and the Hessian of the Lagrangian with respect to* $\mathbf{x}$ *is positive definite on the subspace tangent to* $\mathbf{h}(\mathbf{x})$ *at* $\mathbf{x}_*$*, then* $\mathbf{x}_*$ *is a local constrained minimum.*

Generalized Reduced Gradient (GRG) Method

- Applied the optimality conditions in the reduced space with direct solutions may be impossible and local explorations are needed.
- In the unconstrained case we assume the next iteration is still in the feasible space.
- With equality, any movement in the d-space must be accompanied by adjustments in the s-space
- A decent direction as $\mathbf{d}_k = \mathbf{d}_k - \alpha_k \left(\partial z / \partial \mathbf{d}\right)_k^T$ will have corresponding state variable

$$\begin{aligned}\mathbf{s}'_{k+1} &= \mathbf{s}_k - (\partial \mathbf{h}/\partial \mathbf{s})_k^{-1}(\partial \mathbf{h}/\partial \mathbf{d})_k \partial \mathbf{d}_k \\ &= \mathbf{s}_k + \alpha_k (\partial \mathbf{h}/\partial \mathbf{s})_k^{-1}(\partial \mathbf{h}/\partial \mathbf{d})_k (\partial z/\partial \mathbf{d})_k^T.\end{aligned}$$

Iterative Solutions for Nonlinear Equalities

- The previous state space estimation is based on the linearization of the equality and will not satisfy the equality in general.
- A solution to the nonlinear system

$$\mathbf{h}(\mathbf{d}_{k+1}, \mathbf{s}_{k+1}) = 0$$

give $\mathbf{d}_{k+1}$ can be found iteratively using $\mathbf{s}'_{k+1}$ as a starting point

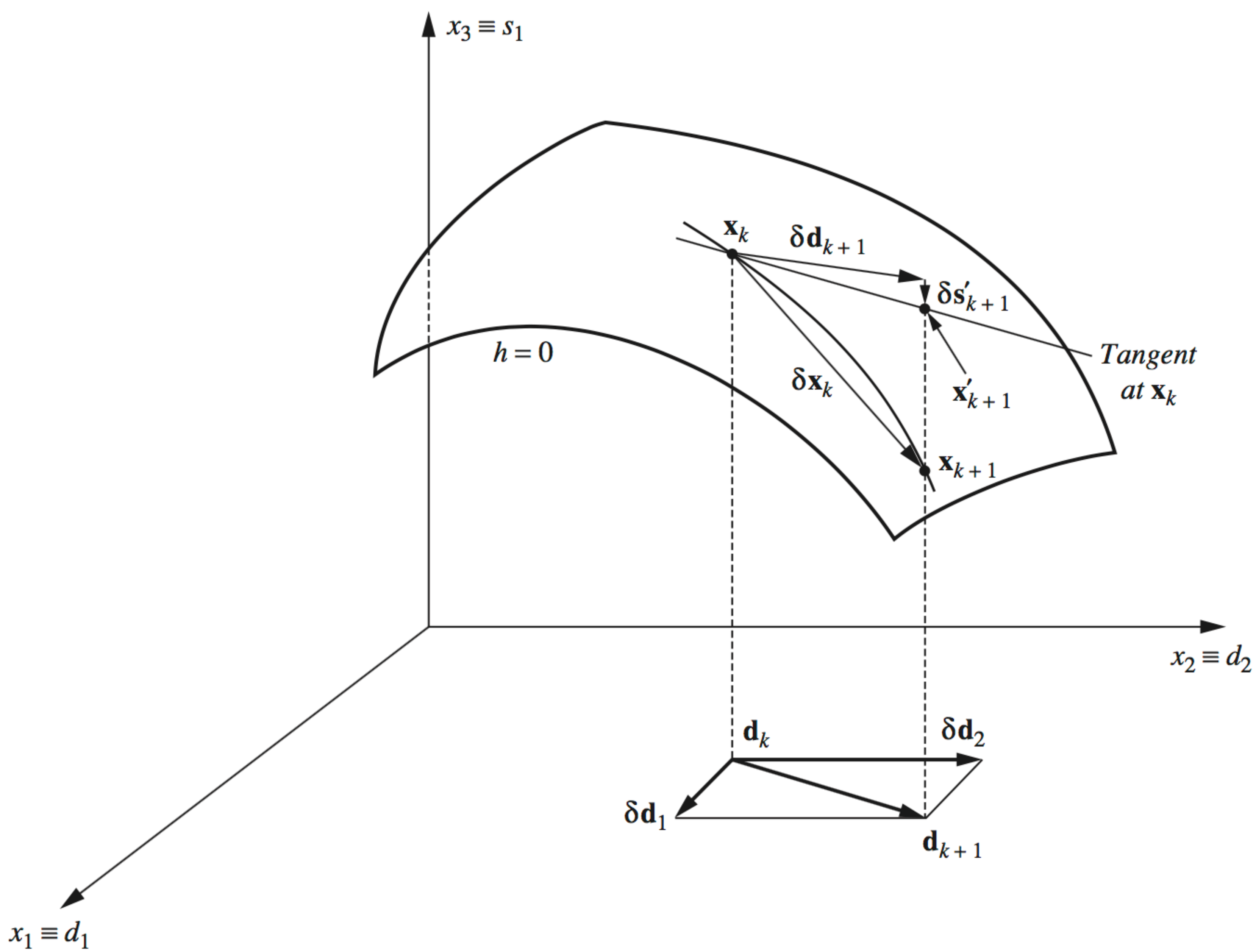
- Therefore we have another inner iteration

$$[\mathbf{s}_{k+1}]_{j+1} = [\mathbf{s}_{k+1} - (\partial\mathbf{h}/\partial\mathbf{s})^{-1}_{k+1}\mathbf{h}(\mathbf{d}_{k+1}, \mathbf{s}_{k+1})]_j,$$

where $\mathbf{s}_{k+1} = \mathbf{s}'_{k+1}$ for j=0.

ME7129
K.-Y. Chan

Concept of Generalized Reduced Gradient



Dealing with Inequalities

$$\min f(\mathbf{x})$$
$$\text{subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}.$$

- Consider problems with inequality. Assume all equalities are removed explicitly or implicitly.
- Only active inequality will be significant.
- Once activities are identified, no difference between active inequality and equality constraints.

$$\nabla f(\mathbf{x}_*) + \boldsymbol{\mu}^T \nabla \overline{\mathbf{g}}(\mathbf{x}_*) = \mathbf{0}^T.$$

Signs of Lagrange Multiplier for Inequality

- The first order perturbation at the optimum must satisfy

$$\begin{aligned} \partial f_* &= \nabla f_* \partial \mathbf{x}_* \geq 0 \quad \text{(optimality)}, \\ \partial \overline{\mathbf{g}}_* &= \nabla \overline{\mathbf{g}}_* \partial \mathbf{x}_* \leq \mathbf{0} \quad \text{(feasibility)}. \end{aligned}$$

- Combined with $\nabla f(\mathbf{x}_*) + \boldsymbol{\mu}^T \nabla \overline{\mathbf{g}}(\mathbf{x}_*) = \mathbf{0}^T.$ the multiplier for inequalities must be strictly nonnegative

- Therefore for all inequalities, we have

$$\nabla f_* + \boldsymbol{\mu}^T \nabla \mathbf{g}_* = \mathbf{0}^T, \quad \underline{\boldsymbol{\mu}^T \mathbf{g} = 0}, \quad \boldsymbol{\mu} \geq \mathbf{0}.$$

complimentary slackness

Karush-Kuhn-Tucker Conditions

- For the general constrained problem

$$\min f(\mathbf{x})$$
$$\text{subject to } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0},$$

- The necessary conditions for optimality (KKT conditions) state

1. $\mathbf{h}(\mathbf{x}_*) = \mathbf{0}, \mathbf{g}(\mathbf{x}_*) \leq \mathbf{0};$
2. $\nabla f_* + \boldsymbol{\lambda}^T \nabla \mathbf{h}_* + \boldsymbol{\mu}^T \nabla \mathbf{g}_* = \mathbf{0}^T, \quad \text{where } \boldsymbol{\lambda} \neq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\mu}^T \mathbf{g} = 0.$

Lagrangian Equations

- In negative null forms

$$\text{minimize} \quad f$$
$$\text{subject to} \quad \mathbf{h} = 0, \ \ \mathbf{g} \leq 0$$

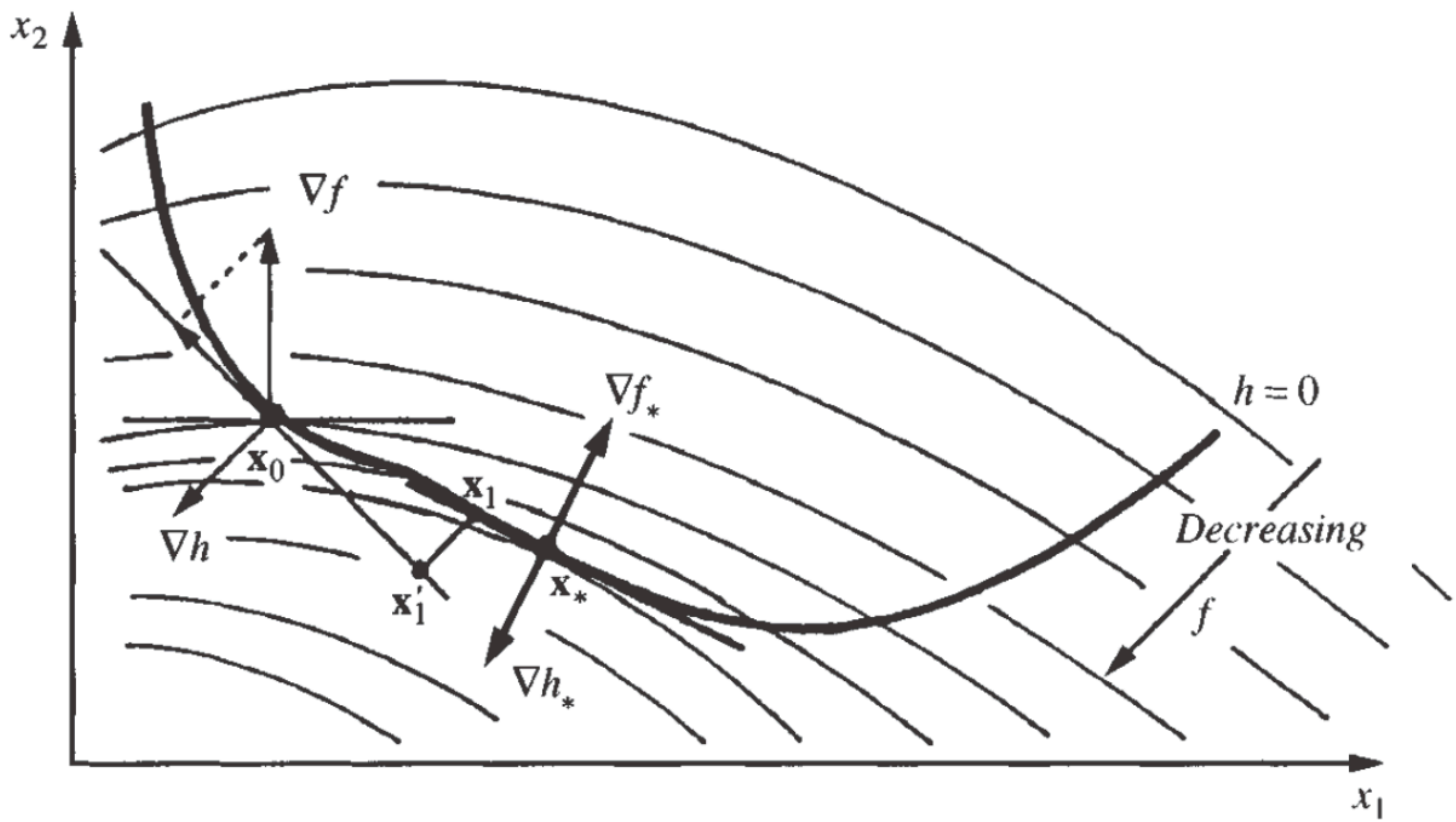
- The Lagrangian is

$$L = f + \boldsymbol{\lambda}^T \mathbf{h} + \boldsymbol{\mu}^T \mathbf{g}$$

- KKT conditions are

$$\nabla f + \boldsymbol{\lambda}^T \nabla \mathbf{h} + \boldsymbol{\mu}^T \nabla \mathbf{g} = \mathbf{0}^T$$
$$\mathbf{h} = 0, \mathbf{g} \leq 0$$
$$\boldsymbol{\lambda} \neq \mathbf{0}, \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\mu}^T \mathbf{g} = 0$$

Geometric Representation of KKT



Linear Programming (LP)

- A problem having objective and constraint functions that are all linear is called a linear programming (LP) problem.

$$\min_{\mathbf{x}} f = \mathbf{c}^T\mathbf{x}$$
$$\text{subject to} \quad \mathbf{h} = \mathbf{A}_1\mathbf{x} - \mathbf{b}_1 = \mathbf{0}$$
$$\mathbf{g} = \mathbf{A}_2\mathbf{x} - \mathbf{b}_2 \leq \mathbf{0}$$

- Standard LP problems can be solved using Simplex Method.

Example 5.12

$$\min_{\mathbf{x}} f = -2x_1 - x_2$$

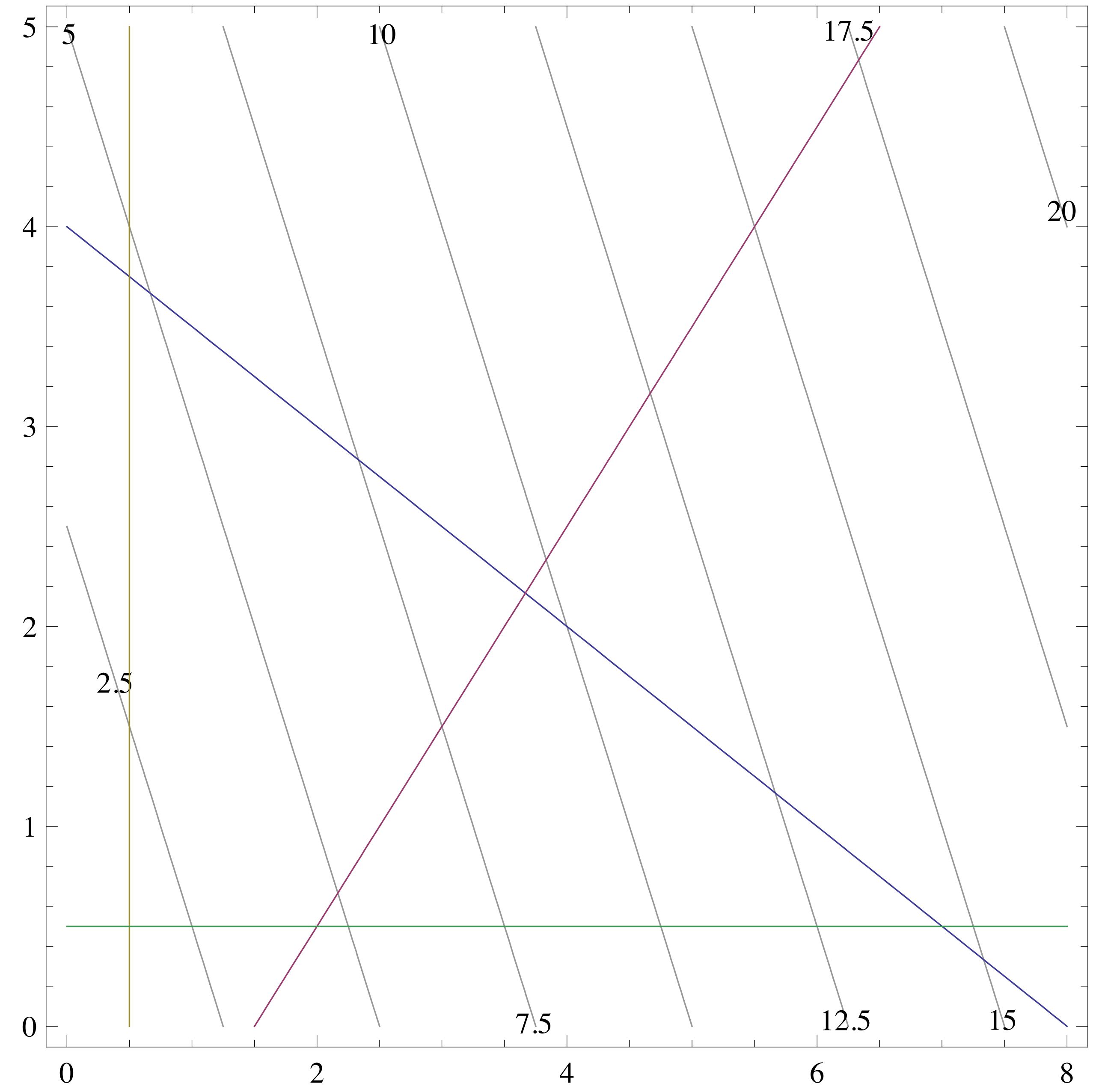
$$\text{subject to} \quad g_1 = x_1 + 2x_2 \leq 8$$

$$g_2 = x_1 - x_2 \leq 3/2$$

$$g_3 = 2x_1 \geq 1$$

$$g_4 = 2x_2 \geq 1$$

	f	g1	g2	g3	g4
x1	-	+	+	-	
x2	-	+	-		-



Optimum of LP

- In a linear model, the objective function and constraints are always monotonic.
- After active (critical) inequality constraints are identified they become equality. Direct eliminations can be used to generate a reduced problem, which will remain monotonic.
- The process will continue as long as activity can be proven, until no variables remain in the objective.
- The solution reached will be at a vertex of the feasible space, which is the intersection of as many active constraint surface as there are variables.

Optimality Conditions

The optimality conditions of the LP problem

$$\begin{aligned} &\min_{\mathbf{x}} f = \mathbf{c}^T\mathbf{x} \\ &\text{subject to} \quad \mathbf{h} = \mathbf{A}_1\mathbf{x} - \mathbf{b}_1 = \mathbf{0} \\ &\qquad\qquad\quad\;\; \mathbf{g} = \mathbf{A}_2\mathbf{x} - \mathbf{b}_2 \leq \mathbf{0} \end{aligned}$$

are

$$\begin{aligned} &\mathbf{A}_1\mathbf{x}_* = \mathbf{b}_1, \mathbf{A}_2\mathbf{x}_* \leq \mathbf{b}_2 \\ &\mathbf{c}^T + \boldsymbol{\lambda}^T\mathbf{A}_1 + \boldsymbol{\mu}^T\mathbf{A}_2 = \mathbf{0}^T, \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\lambda} \neq \mathbf{0} \\ &\boldsymbol{\mu}(\mathbf{A}_2\mathbf{x}_* - \mathbf{b}_2) = 0 \end{aligned}$$

Duality

- Associated with every linear programming problem, hereafter referred to as primal problem, there is a corresponding ***dual*** linear programming problem.
- If one is a maximization problem, the other is a minimization problem.
- The KKT conditions of the primal and dual problems are essentially the same, ***except that the optimality conditions of one are the feasibility conditions of the other and vice versa***.

Dual Problems

Primal

$$\min_{\mathbf{x}} f = \mathbf{c}^T\mathbf{x}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x} + \mathbf{y} = \mathbf{b}$$

$$\mathbf{x} \geq 0, \ \ \mathbf{y} \geq 0$$

Dual

$$\max_{\boldsymbol{\lambda}} f' = \mathbf{b}^T\boldsymbol{\lambda}$$

$$\text{subject to} \quad \mathbf{A}^T\boldsymbol{\lambda} \geq \mathbf{c}$$

$$\boldsymbol{\lambda} \geq 0$$

(1) $\mathbf{A}\mathbf{x} + \mathbf{y} = \mathbf{b}$

(2) $\mathbf{c}^T + \boldsymbol{\lambda}^T\mathbf{A} - \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\mu} \geq 0, \boldsymbol{\lambda} \neq \mathbf{0}$

(3) $\boldsymbol{\mu}\mathbf{x} + \boldsymbol{\lambda}\mathbf{y} = \mathbf{0}$

	Primal	Dual
Feasibility	1	2
Optimality	2	1
Comp. Slackness	3	3