Lagrange Function and KKT Conditions

How do you compute the table of Orthogonal Projections?

$$P_C(\mathbf{z}) = prox_{I_C}(\mathbf{z}) = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2t}||\mathbf{x}-\mathbf{z}||^2 + I_C(\mathbf{x}) = \underset{\mathbf{x}\in C}{\text{argmin}} \frac{1}{2t}||\mathbf{x}-\mathbf{z}||^2$$

Set $C =$	For $t=1$, $P_C(\mathbf{z}) =$	Assumptions
$\Re^n_+$	$[\mathbf{z}]_+$	
$\text{Box}[\mathbf{l},\mathbf{u}]$	$P_C(\mathbf{z})_i = min\{max\{z_i, l_i\}, u_i\}$	$l_i \leq u_i$
$\text{Ball}[\mathbf{c}, r]$	$\mathbf{c} + \frac{r}{max\{\|\mathbf{z}-\mathbf{c}\|_2, r\}}(\mathbf{z}-\mathbf{c})$	$\|.\|_2$ ball, centre $\mathbf{c} \in \Re^n$ & radius $r > 0$
$\{\mathbf{x}\|A\mathbf{x} = \mathbf{b}\}$	$\mathbf{z} - A^T(AA^T)^{-1}(A\mathbf{z}-\mathbf{b})$	$A \in \Re^{m\times n}$, $\mathbf{b} \in \Re^m$, A is full row rank
$\{\mathbf{x}\|\mathbf{a}^T\mathbf{x} \leq b\}$	$\mathbf{z} - \frac{[\mathbf{a}^T\mathbf{x}-b]_+}{\|\mathbf{a}\|^2}$	$0 \neq \mathbf{a} \in \Re^n$ $b \in \Re$
Δ_n	$[\mathbf{z}-\mu^*\mathbf{e}]_+$ where $\mu^* \in \Re$ satisfies $\mathbf{e}^T[\mathbf{z}-\mu^*\mathbf{e}]_+ = 1$	
$H_{\mathbf{a},b} \cap \text{Box}[\mathbf{l},\mathbf{u}]$	$P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}-\mu^*\mathbf{a})$ where $\mu^* \in \Re$ satisfies $\mathbf{a}^T P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}-\mu^*\mathbf{a}) = b$	$0 \neq \mathbf{a} \in \Re^n$ $b \in \Re$
$H^-{}_{\mathbf{a},b} \cap \text{Box}[\mathbf{l},\mathbf{u}]$	$P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z})$ $\mathbf{a}^T P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}) \leq b$ $P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}-\lambda^*\mathbf{a})$ $\mathbf{a}^T P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}) > b$ where $\lambda^* \in \Re$ satisfies $\mathbf{a}^T P_{\text{Box}[\mathbf{l},\mathbf{u}]}(\mathbf{z}-\lambda^*\mathbf{a}) = b$ & $\lambda^* > 0$	$0 \neq \mathbf{a} \in \Re^n$ $b \in \Re$
$B_{\|.\|_1}[0,\alpha]$	$\mathbf{z}$ $\|z\|_1 \leq \alpha$ $[\mathbf{z}-\lambda^*\mathbf{e}]_+ \odot sign(\mathbf{z})$ $\|z\|_1 > \alpha$ where $\lambda^* > 0$, & $[\mathbf{z}-\lambda^*\mathbf{e}]_+ \odot sign(\mathbf{z}) = \alpha$	$\alpha > 0$

Lagrange Function and Necessary KKT Conditions

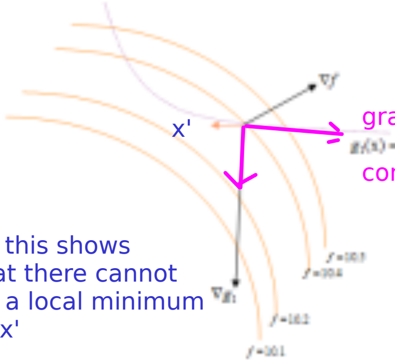
- Can the Lagrange Multiplier construction be generalized to always find optimal solutions to a minimization problem?
- Instead of the iterative path again, assume everything can be computed analytically
- Attributed to the mathematician Lagrange (born in 1736 in Turin). Largely worked on mechanics, the calculus of variations, probability, group theory, and number theory.
- Credited with the choice of base 10 for the metric system (rather than 12).

Lagrange Function and Necessary KKT Conditions

Note that a lot of the analysis that follows does not even assume convexity
Necessary conditions often do NOT require convexity

- Consider the equality constrained minimization problem (with $\mathcal{D} \subseteq \Re^n$)

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{67}$$



x'

grad f has a non-zero component perpendicular to gradient of g1

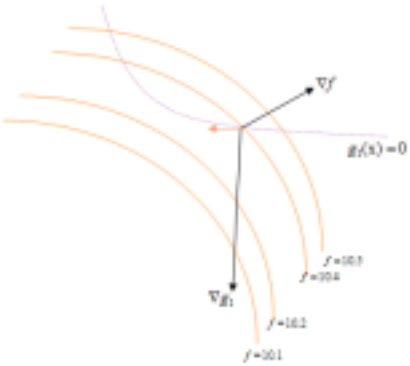
All this shows that there cannot be a local minimum at x'

- The figure shows some level curves of the function f and of a single constraint function g_1 (dotted lines)
- The gradient of the constraint ∇g_1 is not parallel to the gradient ∇f of the function at $f = 10.4$; it is therefore possible to reduce the value of f by moving in negative of non-zero component perpendicular to grad g1

Moving perpendicular to grad g1 ==> g1(x) = 0 remains

Goal: We should not be able to reduce the value of f while still honoring g1(x) = 0

Lagrange Function and Necessary KKT Conditions

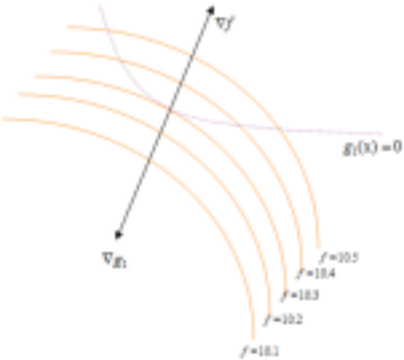


- Consider the equality constrained minimization problem (with $\mathcal{D} \subseteq \Re^n$)

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{67}$$

- The figure shows some level curves of the function f and of a single constraint function g_1 (dotted lines)
- The gradient of the constraint ∇g_1 is not parallel to the gradient ∇f of the function at $f = 10.4$; it is therefore possible to move along the constraint surface so as to further reduce f.

Lagrange Function and Necessary KKT Conditions



- However, ∇g_1 and ∇f are parallel at $f = 10.3$, and any motion along $g_1(\mathbf{x}) = 0$ will

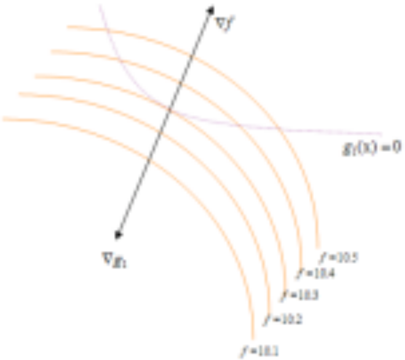
lie along the perpendicular to gradient of g1(x) at that point <==> but gradient of f along that direction = 0!!

==> If we try to decrease value of f, we will land up increasing/decreasing g1 (unacceptable)

==> If we move along perpendicular to gradient of g1, no change expected in f

SO gradients of f and g being in same/opposite directions is necessary condition for local minimum/maximum

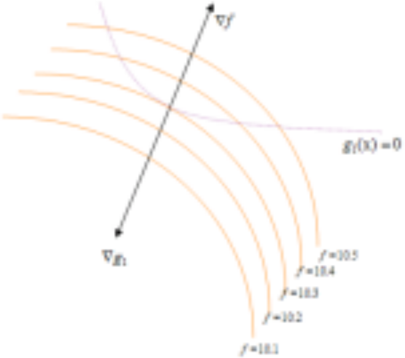
Lagrange Function and Necessary KKT Conditions



- However, ∇g_1 and ∇f are parallel at $f = 10.3$, and any motion along $g_1(\mathbf{x}) = 0$ will leave f unchanged.
- Hence, at the solution $\mathbf{x}^*$,

gradient f(x*) proportional to gradient g1(x*)

Lagrange Function and Necessary KKT Conditions



- However, ∇g_1 and ∇f are parallel at $f = 10.3$, and any motion along $g_1(\mathbf{x}) = 0$ will leave f unchanged.
- Hence, at the solution $\mathbf{x}^*$, $\nabla f(\mathbf{x}^*)$ must be proportional to $-\nabla g_1(\mathbf{x}^*)$, yielding, $\nabla f(\mathbf{x}^*) = -\lambda \nabla g_1(\mathbf{x}^*)$, for some constant $\lambda \in \Re$; λ is called a *Lagrange multiplier.*
- Often λ itself need never be computed and therefore often qualified as the *undetermined* lagrange multiplier.

Lagrange Function and Necessary KKT Conditions

- The necessary condition for an optimum at $\mathbf{x}^*$ for the optimization problem in (68) with $m = 1$ can be stated as in (68); the gradient is now in

The gradient of the Lagrange function wrt x* and lambda* should vanish
as a necessary condition for optimum at x*,lambda*

Lagrange Function and Necessary KKT Conditions

- The necessary condition for an optimum at $\mathbf{x}^*$ for the optimization problem in (68) with $m = 1$ can be stated as in (68); the gradient is now in $\Re^{n+1}$ with its last component being a partial derivative with respect to λ.

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla g_1(\mathbf{x}^*) = 0 \qquad (68)$$
$$g_i(\mathbf{x}^*) = 0$$

- The solutions to (68) are the stationary points of the lagrangian L; they are not necessarily local extrema of L.
  - L is unbounded: given a point $\mathbf{x}$ that doesn't lie on the constraint, letting $\lambda \to \pm\infty$ makes L arbitrarily large or small. (General property of linear functions - here linearity in lambda)
  - However, under certain stronger assumptions, if the *strong Lagrangian principle* holds, the minima of f minimize the Lagrangian globally. A bit later

Lagrange Function and Necessary KKT Conditions

- Let us extend the necessary condition for optimality of a minimization problem with single constraint to minimization problems with multiple equality constraints (*i.e.*, $m > 1$. in (67)).
- Let $\mathcal{S}$ be the subspace spanned by $\nabla g_i(\mathbf{x})$ at any point $\mathbf{x}$ and let $\mathcal{S}_\perp$ be its orthogonal complement. Let $(\nabla f)_\perp$ be the component of ∇f in the subspace $\mathcal{S}_\perp$.

Moving perpendicular to S ==> all constraints remain satisified.
===> At an optimal point x*, we should not be able to move perpendicular to S while reducing the value of f
===> Gradient of cannot have any component along perpendicular to S
===> f MUST lie in S

Lagrange Function and Necessary KKT Conditions

- Let us extend the necessary condition for optimality of a minimization problem with single constraint to minimization problems with multiple equality constraints (*i.e.*, $m > 1$. in (67)).
- Let $\mathcal{S}$ be the subspace spanned by $\nabla g_i(\mathbf{x})$ at any point $\mathbf{x}$ and let $\mathcal{S}_\perp$ be its orthogonal complement. Let $(\nabla f)_\perp$ be the component of ∇f in the subspace $\mathcal{S}_\perp$.
- At any solution $\mathbf{x}^*$, it must be true that the gradient of f has $(\nabla f)_\perp = 0$ (*i.e.*, no components that are perpendicular to all of the ∇g_i), because otherwise you could move $\mathbf{x}^*$ a little in that direction (or in the opposite direction) to increase (decrease) f without changing any of the g_i, *i.e.* without violating any constraints.
- Hence for multiple equality constraints, it must be true that at the solution $\mathbf{x}^*$, the space $\mathcal{S}$ contains the vector ∇f, *i.e.*, there are some constants λ_i such that $\nabla f(\mathbf{x}^*) = \lambda_i \nabla g_i(\mathbf{x}^*)$.

Lagrange Multipliers with Inequality Constraints

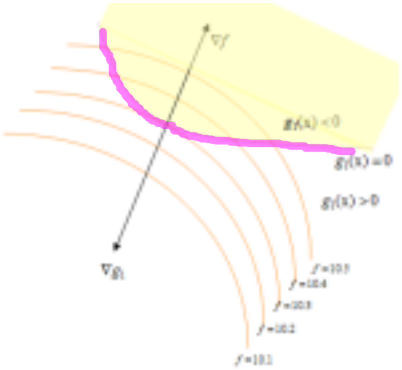
- We also need to impose that the solution is on the correct constraint surface (*i.e.*, $g_i = 0, \ \forall i$). In the same manner as in the case of $m = 1$, this can be encapsulated by introducing the Lagrangian $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$, whose gradient with respect to both $\mathbf{x}$, and λ vanishes at the solution.
- This gives us the following necessary condition for optimality of (67):

Lagrange Multipliers with Inequality Constraints

- We also need to impose that the solution is on the correct constraint surface (*i.e.*, $g_i = 0, \ \forall i$). In the same manner as in the case of $m = 1$, this can be encapsulated by introducing the Lagrangian $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$, whose gradient with respect to both $\mathbf{x}$, and λ vanishes at the solution.
- This gives us the following necessary condition for optimality of (67):

$$\boxed{\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0} \tag{69}$$

Lagrange Multipliers with Inequality Constraints



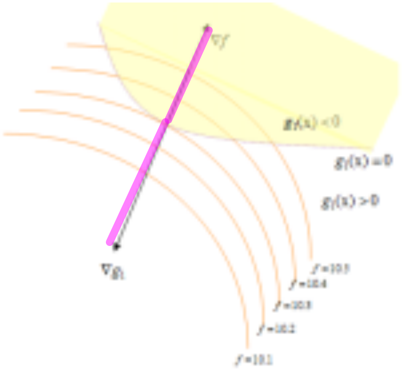
- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active, we must have

gradient of f(x*) and gradient of g(x*) are in same space..

(active case is exactly the same as that of equality constrained optimization)

INACTIVE CONSTRAINT ==> g1(x*) < 0

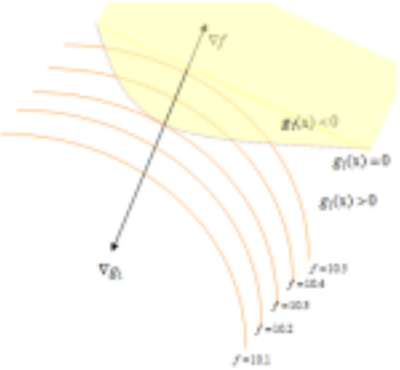
Lagrange Multipliers with Inequality Constraints



- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active, we must have (as in the case of a single equality constraint) that ∇f is parallel to ∇g_1, by the same argument as before.
- Additionally, necessary for the two gradients to point in opposite directions

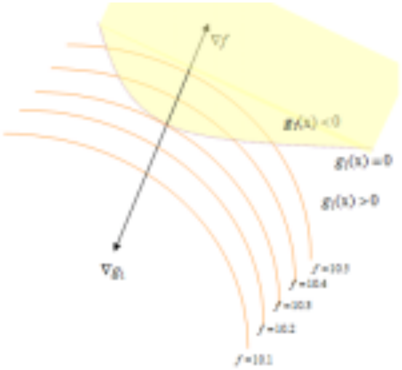
We have a problem: It is fine to reduce f while reducing g1
==> It is fine to move in negative gradient f(x*) if that
also has a component in negative gradient g1(x*)

Lagrange Multipliers with Inequality Constraints



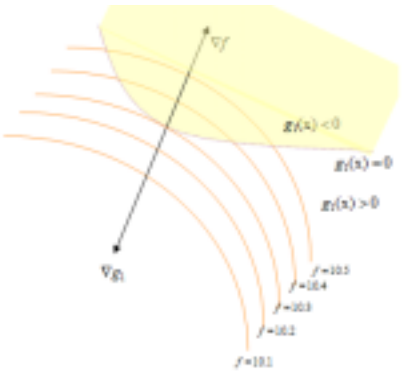
- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active, we must have (as in the case of a single equality constraint) that ∇f is parallel to ∇g_1, by the same argument as before.
- Additionally, necessary for the two gradients to point in opposite directions; else a move away from the surface $g_1 = 0$ and into the feasible region would further reduce f.
- With Lagrangian $L = f + \lambda g_1$, an additional constraint is that lambda1 >= 0

Lagrange Multipliers with Inequality Constraints



- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active, we must have (as in the case of a single equality constraint) that ∇f is parallel to ∇g_1, by the same argument as before.
- Additionally, necessary for the two gradients to point in opposite directions; else a move away from the surface $g_1 = 0$ and into the feasible region would further reduce f.
- With Lagrangian $L = f + \lambda g_1$, an additional constraint is that $\lambda \geq 0$

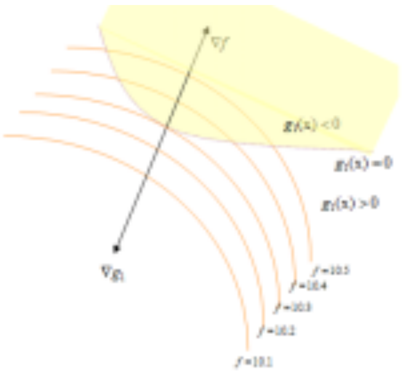
Lagrange Multipliers with Inequality Constraints



- If the constraint is not active at the solution $\nabla f(\mathbf{x}^*) = 0$, then removing g_1 (that is setting lambda1 = 0)

  does not involve lambda1

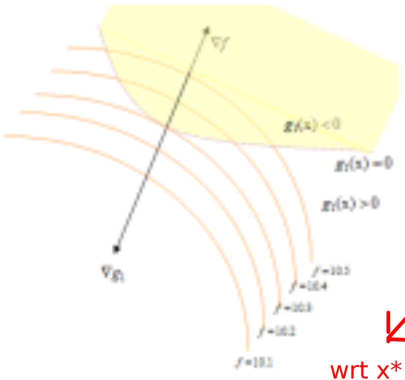
Lagrange Multipliers with Inequality Constraints



- If the constraint is not active at the solution $\nabla f(\mathbf{x}^*) = 0$, then removing g_1 makes no difference and we can drop it from $L = f + \lambda g_1$,
- This is equivalent to setting

lambda1 = 0

Lagrange Multipliers with Inequality Constraints

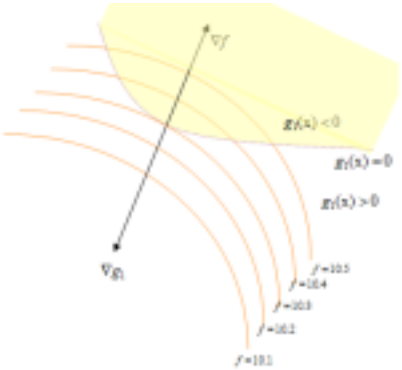


wrt x* only

- If the constraint is not active at the solution $\nabla f(\mathbf{x}^*) = 0$, then removing g_1 makes no difference and we can drop it from $L = f + \lambda g_1$,
- This is equivalent to setting $\lambda = 0$.
- Thus, whether or not the constraints $g_1 = 0$ are active, we can find the solution by requiring that
  1. the gradients of the Lagrangian vanish, and
  2. $\lambda g_1(\mathbf{x}^*) = 0$. (complementary slackness)

  This latter condition is one of the important Karush-Kuhn-Tucker conditions of convex optimization theory that can facilitate the search for the solution and will be more formally discussed subsequently.

Lagrange Multipliers with Inequality Constraints

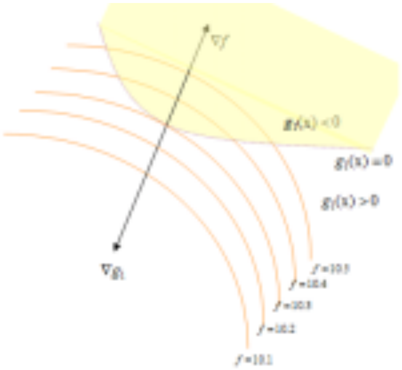


- Now consider the general inequality constrained minimization problem

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{D}} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{70}$$

- With multiple inequality constraints, for constraints that are active, (as in the case of multiple equality constraints),
  1. ∇f must lie in the space spanned by the ∇g_i's,
  2. if the Lagrangian is $L = f + \sum_{i=1}^{m} \lambda_i g_i$, then we must also have $\lambda_i \geq 0$, $\forall i$ (since otherwise f could be reduced by moving into the feasible region).

Lagrange Multipliers with Inequality Constraints



Gradeint is wrt x* only

- As for an inactive constraint g_j ($g_j < 0$), removing g_j from L makes no difference and we can drop ∇g_j from $\nabla f = -\sum_{i=1}^{m} \lambda_i \nabla g_i$ or equivalently set $\lambda_j = 0$.
- Thus, the foregoing KKT condition generalizes to $\lambda_i g_i(\mathbf{x}^*) = 0,\ \forall i$.
- The necessary condition for optimality of (74) is summarized as:

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0$$

$$\forall i \quad \lambda_i g_i(\mathbf{x}) = 0 \quad (71)$$

A simple and often useful trick called the *free constraint gambit* is to solve ignoring one or more of the constraints, and then check that the solution satisfies those constraints, in which case you have solved the problem.

Eg: Take g1 and see if gradient f(x*) + lambda1* gradient g1(x*) = 0 for some lambda1* and x*
If yes, then we have satisified the necessary condition as discussed on the board

A simple and often useful trick called the *free constraint gambit* is to solve ignoring one or more of the constraints, and then check that the solution satisfies those constraints, in which case you have solved the problem.

Some Algebraic Justification: Lagrange Multipliers with Inequality Constraints

Algebraic Justification: Lagrange Multipliers with Inequality Constraints

- For the constrained optimization problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{C} \end{aligned} \tag{72}$$

$$\mathbf{x}^* = \operatorname*{argmin}_{\mathbf{x} \in C} f(\mathbf{x}) \iff \operatorname*{argmin}_{\mathbf{x}} f(\mathbf{x}) + I_C(\mathbf{x}), \text{ where } I_C(\mathbf{x}) = I\{\mathbf{x} \in C\} = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{if } \mathbf{x} \notin C \end{cases}$$

$$N_C(\mathbf{x}) = \partial I_C(\mathbf{x}) = \left\{\mathbf{h} \in \Re^n \,\middle|\, \mathbf{h}^T\mathbf{x} \geq \mathbf{h}^T\mathbf{z} \text{ for any } \mathbf{z} \in C\right\} = \left\{\mathbf{h} \in \Re^n \,\middle|\, \mathbf{h}^T(\mathbf{x} - \mathbf{z}) \geq 0 \text{ for an}\right.$$

- **Recap:** Necessary condition for optimality at $\mathbf{x}^*$: $0 \in \left\{\mathbf{x}^* \,\middle|\, \nabla f(\mathbf{x}^*) + N_{\mathcal{C}}(\mathbf{x}^*)\right\}$, that is, $\nabla f(\mathbf{x}^*) = -N_{\mathcal{C}}(\mathbf{x}^*) = 0$ and therefore

$$\nabla^T f(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) \geq 0 \quad \text{for any } \mathbf{z} \in C \tag{73}$$

Algebraic Justification: Lagrange Multipliers with Inequality Constraints(contd.)

- Specifically, let $C = \left\{\mathbf{x} \in \Re^n \,\middle|\, g_i(\mathbf{x}) \leq 0 \,\forall\, i = 1, 2, \ldots, m\right\}$

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{74}$$

Assume that each g_i is convex and is differentiable. Then, we must have, for each i,

$$\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq g_i(\mathbf{z}) \quad \text{for any } \mathbf{z} \in C \tag{75}$$

- Since $g_i(\mathbf{z}) \leq 0$ whenever $\mathbf{z} \in C$,

Algebraic Justification: Lagrange Multipliers with Inequality Constraints(contd.)

- Specifically, let $C = \left\{\mathbf{x} \in \Re^n \middle| g_i(\mathbf{x}) \leq 0 \,\forall\, i = 1, 2, \ldots, m\right\}$

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{74}$$

Assume that each g_i is convex and is differentiable. Then, we must have, for each i,

$$\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq g_i(\mathbf{z}) \quad \text{for any } \mathbf{z} \in C \tag{75}$$

- Since $g_i(\mathbf{z}) \leq 0$ whenever $\mathbf{z} \in C$,

$$\begin{aligned} & \nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq 0 \quad && \text{for any } \mathbf{z} \in C \\ \Rightarrow \; & -\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) - g_i(\mathbf{x}^*) \geq 0 \quad && \text{for any } \mathbf{z} \in C \end{aligned} \tag{76}$$

Algebraic Justification: Lagrange Multipliers with Inequality Constraints(contd.)

- Since any non-negative scalar (such as in (73)) is a linear combination of non-negative scalars (such as in (76)) with non-negative weights, there exists scalar (vector) $\lambda \in \Re^m_+$ such that

$$\nabla^T f(\mathbf{x}^*)(\mathbf{z}-\mathbf{x}^*) = \sum_{i=1}^{m} -\lambda_i \nabla^T g_i(\mathbf{x}^*)(\mathbf{z}-\mathbf{x}^*) - \lambda_i g_i(\mathbf{x}^*) \quad \text{for any } \mathbf{z} \in C \tag{77}$$

sum of lambdai gi = 0 by substituting z = x*

- Since (77) must hold for any $\mathbf{z} \in C$ and since $\mathbf{x}^* \in C$, we should have $\lambda_i g_i(\mathbf{x}^*) = 0$. Since the equality (77) should also continuously hold on the convex set C, we must also have $\nabla f(\mathbf{x}^*) = \sum_{i=1}^{m} -\lambda_i \nabla g_i(\mathbf{x}^*)$, that is $\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) = 0$
- Since any equality constraint $h_j(\mathbf{x}) = 0$ can be expressed as two inequality constraints: $h_j(\mathbf{x}) \geq 0$ and $-h_j(\mathbf{x}) \geq 0$, the corresponding lagrange multiplier μ_j will have no sign constraints.