RECAP

Demystifying the Projection Step

$$\mathbf{x}_p^{(k+1)} = P_{\mathcal{C}}(\mathbf{x}_u^{(k+1)}) \qquad = \operatorname*{argmin}_{\mathbf{z}} \left\| \mathbf{x}_u^{(k+1)} - \mathbf{z} \right\|_2^2 + I_{\mathcal{C}}(\mathbf{z})$$

$$= \operatorname*{argmin}_{\mathbf{z} \in \mathcal{C}} \left\| \mathbf{x}_u^{(k+1)} - \mathbf{z} \right\|_2^2 \qquad = \operatorname*{argmin}_{\mathbf{z} \in \mathcal{C}} \frac{1}{2} \left\| \mathbf{x}_u^{(k+1)} - \mathbf{z} \right\|_2^2$$

Easy to Project Sets $\mathcal{C}$ (with closed form solutions) RECAP

- Solution set of a linear system $\mathcal{C} = \{\mathbf{x} \in \Re^n : A^T\mathbf{x} = \mathbf{b}\}$
- Affine images $\mathcal{C} = \{A\mathbf{x} + \mathbf{b} : \mathbf{x} \in \Re^n\}$
- Nonnegative orthant $\mathcal{C} = \{\mathbf{x} \in \Re^n : \mathbf{x} \succeq 0\}$. It may be hard to project on arbitrary polyhedron.
- Norm balls $\mathcal{C} = \{\mathbf{x} \in \Re^n : \|\mathbf{x}\|_p \leq 1\}$, for $p = 1, 2, \infty$

Projected Gradient Descent for Affine Constraint Set $\mathcal{C}$ RECAP

Solution set of a linear system $\mathcal{C} = \{\mathbf{x} \in \Re^n : A^T\mathbf{x} = \mathbf{b}\}$

$$\mathbf{x}_p^{(k+1)} = P_{\mathcal{C}}(\mathbf{x}_u^{(k+1)}) = \arg\min_{A^T\mathbf{z}=\mathbf{b}} \frac{1}{2}\left\|\mathbf{x}_u^{(k+1)} - \mathbf{z}\right\|_2^2$$

For $\mathbf{z}, \mathbf{x} \in \Re^n$, A as an $n \times m$ matrix, $\mathbf{b}$ is a vector of size m, consider the slightly more general problem (58) with B as an $n \times n$ matrix:

$$\begin{aligned} \min_{\mathbf{z}\in\Re^n} \quad & \tfrac{1}{2}(\mathbf{z}-\mathbf{x})^T B(\mathbf{z}-\mathbf{x}) \\ \text{subject to} \quad & A^T\mathbf{z} = \mathbf{b} \end{aligned} \tag{58}$$

For projected gradient descent, $B =$

Projected Gradient Descent for Affine Constraint Set $\mathcal{C}$ RECAP

Solution set of a linear system $\mathcal{C} = \{\mathbf{x} \in \Re^n : A^T\mathbf{x} = \mathbf{b}\}$

$$\mathbf{x}_p^{(k+1)} = P_{\mathcal{C}}(\mathbf{x}_u^{(k+1)}) = \arg\min_{A^T\mathbf{z}=\mathbf{b}} \frac{1}{2}\left\|\mathbf{x}_u^{(k+1)} - \mathbf{z}\right\|_2^2$$

For $\mathbf{z}, \mathbf{x} \in \Re^n$, A as an $n \times m$ matrix, $\mathbf{b}$ is a vector of size m, consider the slightly more general problem (58) with B as an $n \times n$ matrix:

$$\begin{aligned} \min_{\mathbf{z}\in\Re^n} \quad & \frac{1}{2}(\mathbf{z}-\mathbf{x})^T B(\mathbf{z}-\mathbf{x}) \\ \text{subject to} \quad & A^T\mathbf{z} = \mathbf{b} \end{aligned} \tag{58}$$

For projected gradient descent, $B = I$. Further, if $n = 2$ and $m = 1$, the minimization problem (58) amounts to finding a point $\mathbf{y}^*$ on a line $a_{11}z_1 + a_{12}z_2 = b$ that is closest to $\mathbf{x}$.

Projected Gradient Descent for Affine Constraint Set $\mathcal{C}$ RECAP

- Consider minimization of the modified objective function $L(\mathbf{z}, \lambda) = \frac{1}{2}(\mathbf{z} - \mathbf{x})^T B(\mathbf{z} - \mathbf{x}) + \lambda^T(A^T\mathbf{z} - \mathbf{b})$.

$$\min_{\mathbf{z}\in\Re^n, \lambda\in\Re^m} \quad \frac{1}{2}(\mathbf{z} - \mathbf{x})^T B(\mathbf{z} - \mathbf{x}) + \lambda^T(A^T\mathbf{z} - \mathbf{b}) \tag{59}$$

  The function $L(\mathbf{z}, \lambda)$ is called the lagrangian and involves the lagrange multiplier $\lambda \in \Re^m$.
- A sufficient condition for optimality of $L(\mathbf{z}, \lambda)$ at a point $L(\mathbf{z}^*, \lambda^*)$ is that $\nabla L(\mathbf{z}^*, \lambda^*) = 0$ and $\nabla^2 L(\mathbf{z}^*, \lambda^*) \succ 0$. For this specific problem:

$$\nabla L(\mathbf{z}^*, \lambda^*) = \begin{bmatrix} B\mathbf{z}^* - \frac{1}{2}(B + B^T)\mathbf{x} + A\lambda^* \\ A^T\mathbf{z}^* - \mathbf{b} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  and

$$\nabla^2 L(\mathbf{z}^*, \lambda^*) = \begin{bmatrix} B & A \\ A^T & 0 \end{bmatrix} \succ 0$$

Projected Gradient Descent for Affine Constraint Set $\mathcal{C}$

RECAP

- The point $(\mathbf{z}^*, \lambda^*)$ must therefore satisfy, $A^T\mathbf{z}^* = \mathbf{b}$ and $A\lambda^* = -B\mathbf{z}^* + \frac{1}{2}(B + B^T)\mathbf{x}$.
- Recap: If B is taken to be the identity matrix, $n = 2$ and $m = 1$, the minimization problem (58) amounts to finding a point $\mathbf{y}^*$ on a line $a_{11}z_1 + a_{12}z_2 = b$ that is closest to $\mathbf{x}$.
- From geometry, the point on a line closest to $\mathbf{x}$ is the point of intersection $\mathbf{p}^*$ of a perpendicular (or least possible[8] obtuse angle) from $\mathbf{x}$ to the line. However, the solution for the minimum of (59), for these conditions coincides with $\mathbf{p}^*$ and is given by:

$$z_1^* = x_1 - \frac{a_{11}(a_{11}x_1 + a_{12}x_2 - b)}{(a_{11})^2 + (a_{12})^2} \qquad z_2^* = x_2 - \frac{a_{12}(a_{11}x_1 + a_{12}x_2 - b)}{(a_{11})^2 + (a_{12})^2}$$

That is, for $n = 2$ and $m = 1$, the solution to (59) is the same as the solution to (58)

- For general n and m,

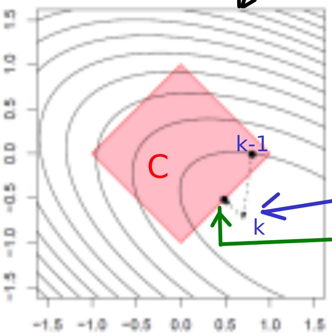
$$\mathbf{z}^* = \mathbf{x}_p^{(k+1)} = P_{\mathcal{C}}(\mathbf{x}_u^{(k+1)}) = \arg\min_{A^T\mathbf{z}=\mathbf{b}} \frac{1}{2}\left\|\mathbf{x}_u^{(k+1)} - \mathbf{z}\right\|_2^2 = \mathbf{x}_u^{(k+1)} - A(A^TA)^{-1}(A^T\mathbf{x}_u^{(k+1)} - \mathbf{b})$$

Was this an accident?
More today!!

[8] See following slides for some elaboration on geometry of the projection

Projected Gradient Descent: Illustrated and Summarized

Level surfaces for the quadratic objective



- Illustration of Projected Gradient Descent on Quadratic Objective with bounded affine (Polyhedral) constraint set
- The line joining point of projection $\mathbf{x}_p^k = P_C(\mathbf{x}_u^k)$ to $\mathbf{x}_u^k$ forms least possible obtuse angle[a] with line joining $\mathbf{x}_p^k = P_C(\mathbf{x}_u^k)$ to any point $\mathbf{z} \in C$.

[a] See following slides for some elaboration on geometry of the projecti

Elaboration on the Geometry of the Projected Gradient Descent

Right angle FOR Affine Set/Unbounded sets

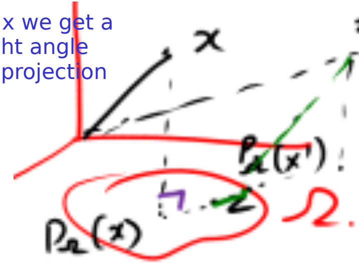
Least possible obtuse angle FOR Polyhedron/Bounded Sets

- **Claim:** If $P_{\mathcal{C}}(\mathbf{x})$ is a projection of $\mathbf{x}$, then

$$(\mathbf{z} - P_{\mathcal{C}}(\mathbf{x}))^{\top}(\mathbf{x} - P_{\mathcal{C}}(\mathbf{x})) \leq 0, \forall\, \mathbf{z} \in \mathcal{C}$$

- That is, the angle between $(z - P_{\mathcal{C}}(x))$ and $(x - P_{\mathcal{C}}(x))$ is obtuse (or right-angled for the projected point), $\forall z \in \mathcal{C}$

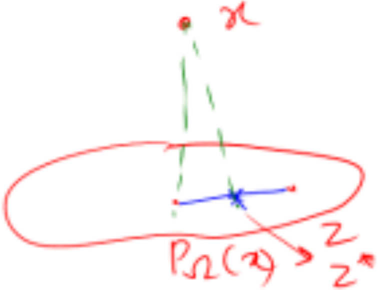
At x we get a
right angle
at projection



At x' we make an obtuse angle at projection

Proof for $\langle z - P_{\mathcal{C}}(x), x - P_{\mathcal{C}}(x) \rangle \leq 0$

- To be more general, let us consider an inner product $\langle a, b \rangle$ instead of $a^\top b$
- Let $z^* = (1-\alpha)P_{\mathcal{C}}(x) + \alpha z$, for some $\alpha \in (0,1)$, and $z \in \mathcal{C}$
  $\implies z^* = P_{\mathcal{C}}(x) + \alpha(z - P_{\mathcal{C}}(x))$, $z^* \in \mathcal{C}$



- Since $P_{\mathcal{C}}(x) = \text{argmin}_{z \in \mathcal{C}} \|x - z\|_2^2$,
  $\left\|x - P_{\mathcal{C}}(x)\right\|^2 \leq \|x - z^*\|^2$

$$
\begin{aligned}
&\|x - z^*\|^2 \\
&= \left\| x - \left(P_{\mathcal{C}}(x) + \alpha(z - P_{\mathcal{C}}(x)) \right) \right\|^2 \\
&= \left\| x - P_{\mathcal{C}}(x) \right\|^2 + \alpha^2 \left\| z - P_{\mathcal{C}}(x) \right\|^2 - 2\alpha \left\langle x - P_{\mathcal{C}}(x), z - P_{\mathcal{C}}(x) \right\rangle \\
&\geq \left\| x - P_{\mathcal{C}}(x) \right\|^2
\end{aligned}
$$

$$
\implies \left\langle x - P_{\mathcal{C}}(x), z - P_{\mathcal{C}}(x) \right\rangle \leq \frac{\alpha}{2} \left\| z - P_{\mathcal{C}}(x) \right\|^2, \ \forall \alpha \in (0, 1)
$$

- Thus, the LHS can either be 0 or a negative value. Any positive value of the LHS will lead to a contradiction for some small $\alpha \to 0$
- Hence, we proved that $\left\langle z - P_{\mathcal{C}}(x), x - P_{\mathcal{C}}(x) \right\rangle \leq 0$

If x* is in the set C, it itself must be the projection

- We can also prove that if $\langle x - x^*, z - x^* \rangle \leq 0$, $\forall z \in \mathcal{C}$ s.t. $z \neq x^*$, and $x^* \in \mathcal{C}$, then

$$x^* = P_{\mathcal{C}}(x) = \underset{\bar{z} \in \mathcal{C}}{\operatorname{argmin}} \|x - \bar{z}\|_2^2$$

- Consider $\|x - z\|^2 - \|x - x^*\|^2$
$= \left\|x - x^* + (x^* - z)\right\|^2 - \|x - x^*\|^2$
$= \|x - x^*\|^2 + \|z - x^*\|^2 - 2\langle x - x^*, z - x^* \rangle - \|x - x^*\|^2$
$= \|z - x^*\|^2 - 2\langle x - x^*, z - x^* \rangle$
> 0
- $\implies \|x - z\|^2 > \|x - x^*\|^2$, $\forall z \in \mathcal{C}$ s.t. $z \neq x^*$
- This proves that $x^* = P_{\mathcal{C}}(x)$

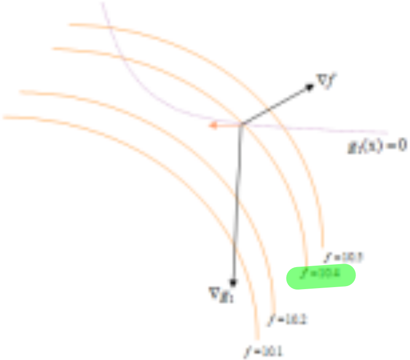
Lagrange Function and KKT Conditions

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. Attributed choice of base 10 for the metric system (rather than 12).

Projected gradient descent is only one consumer for this analysis.
There are several other results and algorithms that make use of this analysis

Lagrange Function and Necessary KKT Conditions

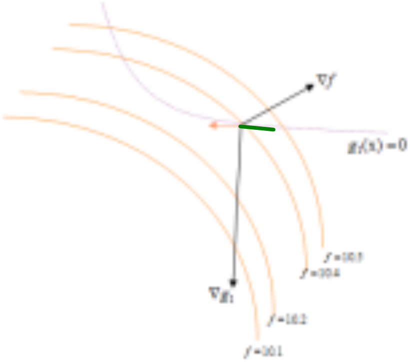


- Consider the equality constrained minimization problem (with $\mathcal{D} \subseteq \Re^n$)

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) = 0 \quad i = 1, 2, \ldots, m \end{aligned} \tag{60}$$

- 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 decrease f while maintaining g1(x) = 0 (by moving tangential to g1(x) = 0

Lagrange Function and Necessary KKT Conditions

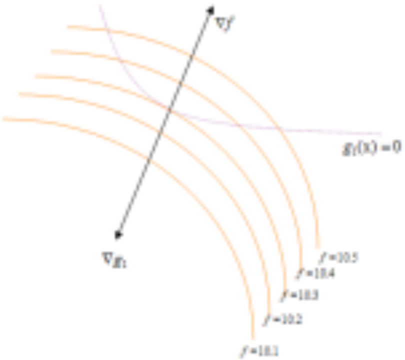


- Consider the equality constrained minimization problem (with $\mathcal{D} \subseteq \Re^n$)

$$\begin{array}{lll} \min\limits_{\mathbf{x}\in\mathcal{D}} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) = 0 & i = 1, 2, \ldots, m \end{array} \tag{60}$$

- 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

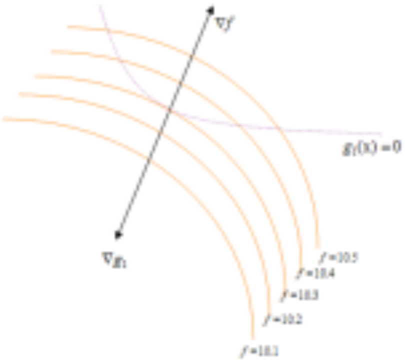


- However, ∇g_1 and ∇f are parallel at $f = 10.3$, and any motion along $g_1(\mathbf{x}) = 0$ will

  not change the value of f(x) since gradient of f has no component perpendicular to the gradient of g1(x) = 0

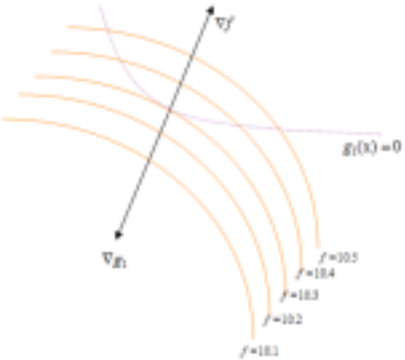
At x s.t f(x) = 10.3,
gradient of f = \lambda gradient of g1
sign of \lambda does not matter

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 increase f, or leave it unchanged.
- Hence, at the solution $\mathbf{x}^*$,

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 increase f, or leave it 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 (60) with $m = 1$ can be stated as in (61); the gradient is now in

Lagrange Function and Necessary KKT Conditions

- The necessary condition for an optimum at $\mathbf{x}^*$ for the optimization problem in (60) with $m = 1$ can be stated as in (61); the gradient is now in $\Re^{n+1}$ with its last component being a partial derivative with respect to λ.

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla f(\mathbf{x}^*) + \lambda^* \nabla g_1(\mathbf{x}^*) = 0 \tag{61}$$

- The solutions to (61) 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. However, under certain stronger assumptions, if the *strong Lagrangian principle* holds, the minima of f minimize the Lagrangian globally.

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 (60)).
- 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 $\mathcal{S}_\perp$.

There is no component of gradient f perpendiculat to S

SAME AS

gradient of f lies in S

Lagrange Function and Necessary KKT Conditions

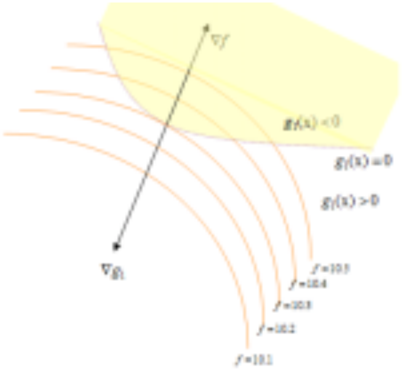
- Let us extend the necessary condition for optimality of a minimization problem with single constraint to minimization problems with multiple equality constraints (*i.e.*, $m > 1$. in (60)).
- Let $\mathcal{S}$ be the subspace spanned by $\nabla g_i(\mathbf{x})$ at any point $\mathbf{x}$ and let $\mathcal{S}_\perp$ 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 (60):

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0 \quad (62)$$

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

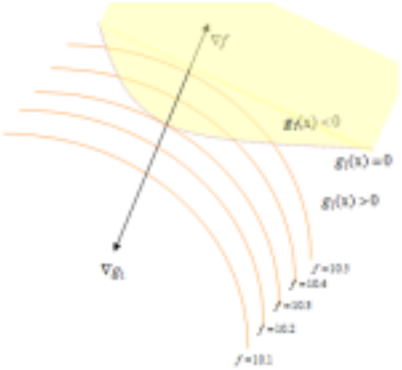
  gradient of f(x*) has no component perpendicular to gradient g1(x*)

  AND

  - gradient of f(x*) is not along direction of - gradient of g1(x*)

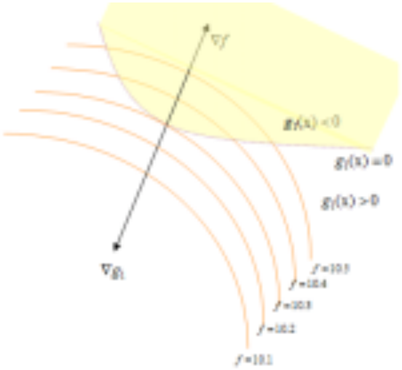
  THAT IS, the two gradients MUST be in opposite directions

Lagrange Multipliers with Inequality Constraints



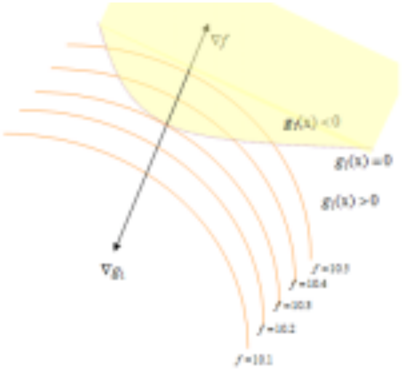
- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active,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

Lagrange Multipliers with Inequality Constraints



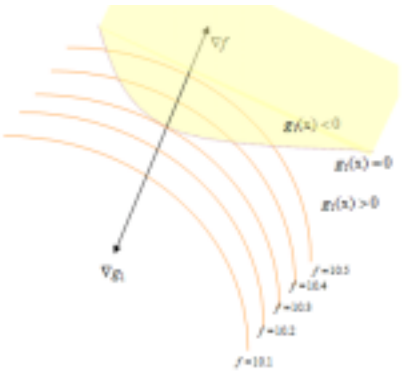
- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active,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

Lagrange Multipliers with Inequality Constraints



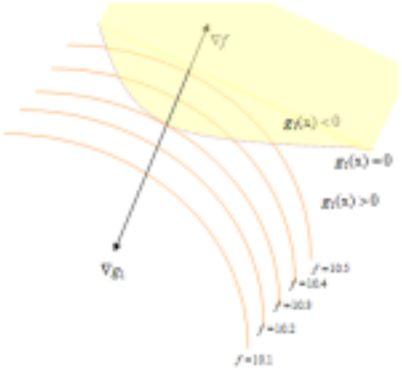
- Single equality constraint $g_1(\mathbf{x}) = 0$, replaced with a single inequality constraint $g_1(\mathbf{x}) \leq 0$. The entire region labeled $g_1(\mathbf{x}) \leq 0$ in the Figure becomes feasible.
- At the solution $\mathbf{x}^*$, if $g_1(\mathbf{x}^*) = 0$, *i.e.*, if the constraint is active,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



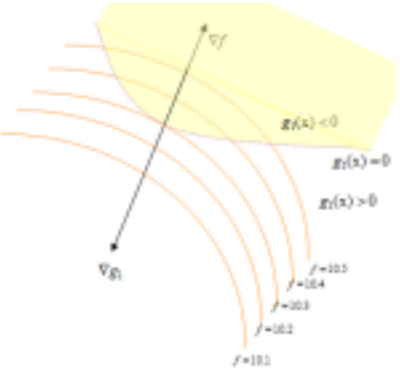
- If the constraint is not active at the solution $\nabla f(\mathbf{x}^*) = 0$, then removing g_1

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

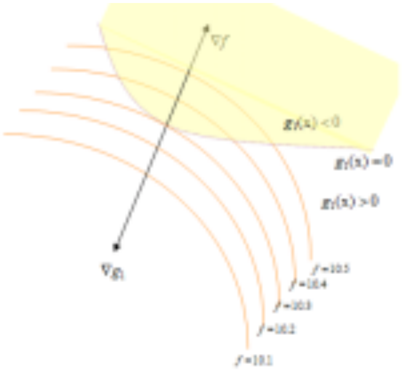
Lagrange Multipliers with Inequality Constraints



- 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 $\lambda = 0$.
- 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. $\lambda g_1(\mathbf{x}^*) = 0$.

  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

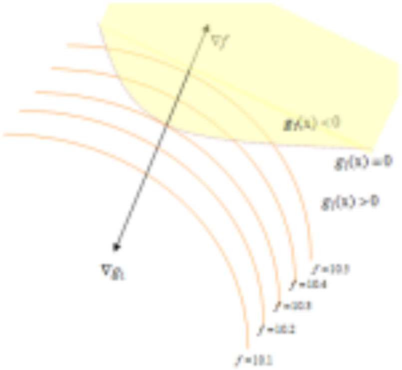


- Now consider the general inequality constrained minimization problem

$$\begin{array}{lll} \min\limits_{\mathbf{x}\in\mathcal{D}} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \le 0 & i = 1, 2, \ldots, m \end{array} \quad (63)$$

- 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 $\lambda_i \ge 0,\ \forall i$ (since otherwise f could be reduced by moving into the feasible region).

Lagrange Multipliers with Inequality Constraints



- 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 (67) is summarized as:

$$\nabla L(\mathbf{x}^*, \lambda^*) = \nabla \left(f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right) = 0$$
$$\forall i \ \ \lambda_i g_i(\mathbf{x}) = 0 \quad (64)$$

- A simple and often useful trick called the free

Some Algebraic Justification: Lagrange Multipliers with Inequality Constraints

Algebraic Justification: Lagrange Multipliers with Inequality Constraints

- For the constrained optimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}\in\mathcal{C} \end{aligned} \tag{65}$$

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

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

- 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

zero belongs to the subdifferential

Negative of gradient of f at x*
must lie in normal cone

$$\nabla^T f(\mathbf{x}^*)(\mathbf{z}-\mathbf{x}^*) \geq 0 \quad \text{for any } \mathbf{z}\in C \tag{66}$$

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

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

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

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

First order condition for convexity of gi

$$\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq g_i(\mathbf{z}) \quad \text{for any } \mathbf{z} \in C \tag{68}$$

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

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

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

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

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

$$\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq g_i(\mathbf{z}) \quad \text{for any } \mathbf{z} \in C \tag{68}$$

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

$$\begin{aligned} & \nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) + g_i(\mathbf{x}^*) \leq 0 && \text{for any } \mathbf{z} \in C \\ \Rightarrow\; & -\nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) - g_i(\mathbf{x}^*) \geq 0 && \text{for any } \mathbf{z} \in C \end{aligned} \tag{69}$$

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

- Since any non-negative scalar (such as in (66)) is a linear combination of non-negative scalars (such as in (69)) with non-negative weights, there exists scalar (vector) $\lambda \in \Re^m_+$ such that

$$\nabla^T f(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) = \sum_{i=1}^{m} -\lambda_i \nabla^T g_i(\mathbf{x}^*)(\mathbf{z} - \mathbf{x}^*) - \lambda_i g_i(\mathbf{x}^*) \quad \text{for any } \mathbf{z} \in C \tag{70}$$

- Since (70) 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 (70) 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. Additionally we require -h and h to be both convex ==> h is affine

Duality Theory for Constrained Optimization

A tricky thing in duality theory is to decide what we call the domain or *ground set* $\mathcal{D}$ and what we call the constraints g_i's or h_j's. Based on whether constraints are explicitly stated or implicitly stated in the form of the ground set, the dual problem could be very different. Thus, many duals are possible for the given primal.

For the rest of the discussion $\mathcal{D}$ will mostly mean $\Re^n$

Formally: The Dual Theory for Constrained Optimization

Consider the general constrained minimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \\ \text{subject to} \quad & h_j(\mathbf{x}) = 0,\ j = 1, 2, \ldots, n \end{aligned} \tag{71}$$

- Consider forming the lagrange function by associating prices (called lagrange multipliers) λ_i and μ_j , with constraints involving g_i and h_j respectively.

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{n} \mu_j h_j(\mathbf{x}) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + \mu^T \mathbf{h}(\mathbf{x})$$

- At each feasible $\mathbf{x}$, for fixed $\lambda_i \geq 0\ \forall i \in \{1..m\}$,

f(x) is lower bounded by the value of the Lagrange function
for all primal feasible x and dual feasible \lambda

Formally: The Dual Theory for Constrained Optimization

Consider the general constrained minimization problem

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{D}} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \\ &\text{subject to} && h_j(\mathbf{x}) = 0,\ j = 1, 2, \ldots, n \end{aligned} \tag{71}$$

- Consider forming the lagrange function by associating prices (called lagrange multipliers) λ_i and μ_j , with constraints involving g_i and h_j respectively.

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{n} \mu_j h_j(\mathbf{x}) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}) + \mu^T \mathbf{h}(\mathbf{x})$$

- At each feasible $\mathbf{x}$, for fixed $\lambda_i \geq 0\, \forall i \in \{1..m\}$,

$$f(\mathbf{x}) \geq \quad L(\mathbf{x}, \lambda, \mu) \quad \text{if } g_i(\mathbf{x}) \leq 0 \ \&\ h_j(\mathbf{x}) = 0 \tag{72}$$

Formally: The Dual Theory for Constrained Optimization

- For $\lambda_i \geq 0 \, \forall i \in \{1..m\}$ and μ_j, minimizing the right hand side of (72) over all feasible $\mathbf{x}$

$$f(\mathbf{x}) \geq \min_{\mathbf{x} \ s.t \ g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0} L(\mathbf{x}, \lambda, \mu) \triangleq L^*(\lambda, \mu) \qquad (73)$$

- $L^*(\lambda, \mu)$ is a pointwise (w.r.t $\mathbf{x} \in g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0$) minimum of linear functions ($L(\mathbf{x}, \lambda, \mu)$) and is therefore always a concave

L(...) for a fixed x is affine function of lambda and mu

RECAP: Pointwise max/supremum of affine functions is always convex

Formally: The Dual Theory for Constrained Optimization

- For $\lambda_i \geq 0\, \forall i \in \{1..m\}$ and μ_j, minimizing the right hand side of (72) over all feasible $\mathbf{x}$

$$f(\mathbf{x}) \geq \min_{\mathbf{x}\ s.t\ g_i(\mathbf{x})\leq 0, h_j(\mathbf{x})=0} L(\mathbf{x}, \lambda, \mu) \quad \triangleq L^*(\lambda, \mu) \tag{73}$$

- $L^*(\lambda, \mu)$ is a pointwise (w.r.t $\mathbf{x} \in g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0$) minimum of linear functions ($L(\mathbf{x}, \lambda, \mu)$) and is therefore always a concave function.
- Since $f(\mathbf{x}) \geq L^*(\lambda, \mu)$ for all primal feasible $\mathbf{x}$ and dual feasible *i.e.*, $\lambda_i \geq 0$ and μ_j, , we can maximize the lower bound $L^*(\lambda, \mu)$ to give the following **Dual Problem**

$$\begin{aligned} \max_{\lambda \in \Re^m, \mu \in \Re^p} \quad & L^*(\lambda, \mu) \\ \text{subject to} \quad & \lambda \geq \mathbf{0} \end{aligned} \tag{74}$$

Theorem

(i) The dual function $L^*(\lambda, \mu)$ is always concave. (ii) If p^* is solution of (71) and d^* of (74) then $p^* \geq d^*$