CISC 371 Class 24

Inequality Constraints and KKT Conditions

Texts: [1] pp. 191–203, 207–222; [2] pp. 397–398; [3] pp. 320–336

Main Concepts:

- *Linear inequality:* $A\vec{w} \leq \vec{b}$
- *Linear inequalities describe a convex set*
- *Feasible point: point satisfying every inequality*
- *KKT conditions: necessary and sufficient for linear inequalities*

Sample Problem, Machine Inference: What is the vector of minimum length that is inside a set with linear boundaries?

The next problem that we will explore is optimization of a quadratic objective with linear equality constraints. This type of problem occurs widely and has an extensive history.

For this course, a linear inequality constraint of a vector $\vec{w}$ will be written as

$$\underline{a}\vec{w} \leq b$$

If there are m linear inequalities, then we can combine the vectors $\vec{a}_i$ into a single $m \times n$ matrix. This represents all of the inequality constraints as the vector inequality

$$A\vec{w} \leq \vec{b} \tag{24.1}$$

For an objective function $f(\vec{w})$, subject to a set of linear inequality constraints of the form in Equation 24.1, the constrained optimization problem can be formulated as a Lagrange function. For the i^{th} constraint we will write the Lagrange multiplier as λ_i. Gathering the Lagrange multipliers into a vector $\vec{\lambda}$, our problems can be stated as a general constrained optimization that can then be simplified to one that has inear inequality constraints.

Problem: quadratic objective, linear inequality constraints

Combining a set of p linear inequalities in the form of Equation 24.1, we can write this problem as

$$\begin{array}{rrcl} & \vec{w}^* & = & \underset{\vec{w}\in\mathbb{R}^n}{\operatorname{argmin}} f(\vec{w}) \\ & \vec{p}(\vec{w}^*) & \leq & \vec{0} \\ \text{where, for } K = K^T \text{ and } A \in \mathbb{R}^{m\times n}, & & & \\ & f(\vec{w}) & = & \frac{1}{2}\vec{w}^T K\vec{w} + \vec{q}^T\vec{w} \\ & \vec{p}(\vec{w}) & = & A\vec{w} - \vec{b} \\ & K & \succeq & 0 \end{array} \tag{24.2}$$

The Lagrange formula for Problem 24.2 is

$$
\begin{aligned}
\mathcal{L}(\vec{w}, \vec{\lambda}) &= f(\vec{w}) + \sum_{i=1}^{m} \lambda_i p_i(\vec{w}) \\
&= f(\vec{w}) + \vec{\lambda} \cdot \vec{p}(\vec{w}) \\
&= f(\vec{w}) + \vec{\lambda}^T \vec{p}(\vec{w}) \qquad (24.3)
\end{aligned}
$$

The corresponding Lagrange equation for Equation 24.3, written using vectors rather than using 1-forms, is

$$
\begin{bmatrix} \left[\dfrac{\partial \mathcal{L}}{\partial \vec{w}}\right]^T \\ \\ \left[\dfrac{\partial \mathcal{L}}{\partial \vec{\lambda}}\right]^T \end{bmatrix} \left(\vec{w}, \vec{\lambda}^*\right) = \vec{0} \qquad (24.4)
$$

24.1 Example: Scalar Argument, Single Inequality

A simple example is a quadratic function $f_1(t)$ that has a scalar argument, and a single inequality constraint that $t \leq -1$:

$$
\begin{aligned}
f_1(t) &= (t-2)(t-4) \qquad (24.5) \\
(1)t &\leq -1
\end{aligned}
$$

The minimizer of $f_1(t)$ is $t^* = +3$, but the minimizer of the constrained problem in Problem 24.5 is $t^* = -1$. A plot of the function and the constraint is shown in Figure 24.1.

When we study the plot in Figure 24.1, we can see that the minimizer occurs when the inequality constraint is satisfied as an equality constraint. We can solve the problem by finding the stationary point of the Lagrange function

$$
\mathcal{L}(t, \lambda) = (t-2)(t-4) + \lambda(t+1) \qquad (24.6)
$$

The stationary point of Equation 24.6 is

$$
t^* = -1 \qquad \lambda^* = +8 \qquad (24.7)
$$

Now, let us retain the convex function $f_1(t)$ and change the single inequality constraint to $t \leq 5$:

$$
\begin{aligned}
f_1(t) &= (t-2)(t-4) \qquad (24.8) \\
(1)t &\leq 5
\end{aligned}
$$

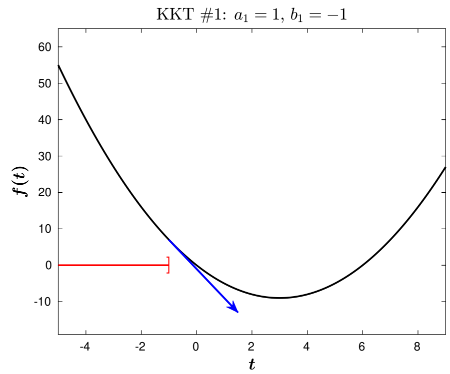


Figure 24.1: The function $f_1(t)$ is shown in black and the constraint interval is shown in red. The minimizer of the constrained problem is $t^* = -1$. At the minimizer, the function $f_1(t)$ is not stationary; the derivative at $f_1(t^*)$, shown in blue, is not horizontal.

The minimizer of $f_1(t)$ is $t^* = +3$ and, because this satisfies the constraint $t^* \leq 5$, this is also the minimizer of the constrained problem in Problem 24.8. A plot of the function and the constraint is shown in Figure 24.2.

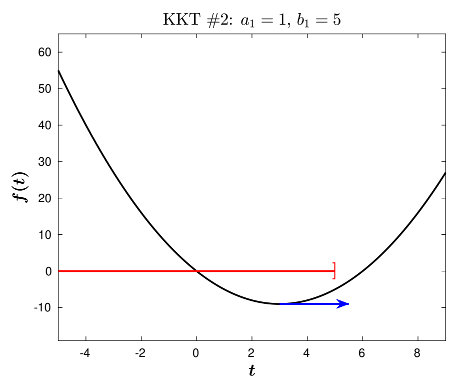


Figure 24.2: The function $f_1(t)$ is shown in black and the constraint interval is shown in red. The minimizer of the constrained problem is $t^* = +3$. At the minimizer, the function $f_1(t)$ is stationary; the derivative at $f_1(t^*)$, shown in blue, is horizontal.

<u>Observation</u>: In Problem 24.5 the constraint $t^* \leq -1$ was *active* and we needed to solve the constrained problem. In Problem 24.8 the constraint $t^* \leq 5$ was *inactive* and we needed to solve the ordinary unconstrained problem.

We can resolve the difference between Problem 24.5 and Problem 24.8 by setting the Lagrange multiplier for Problem 24.8 to be $\lambda^* = 0$. This converts Equation 24.6 to an unconstrained optimization problem, for which $t^* = +3$.

Observation: The Lagrange multiplier of an inactive constraint is zero. For a consistent problem, the Lagrange multiplier of an inactive constraint is non-negative.

Example: Planar Points, Three Active Constraints

For example, suppose that our search space is $\mathbb{R}^2$, so every search vector $\vec{w}$ has two entries. Unlike equality constraints, the matrix in Equation 24.1 does not need to be full rank. It is possible, and sometimes necessary, for $m \gg n$ in some problems.

Recall that a point $\hat{w}$ is a *feasible solution* to a problem if it satisfies the constraints of the problem. For linear inequality constraints, such a point must be such that $p(\hat{w}) \leq 0$, or

$$A\hat{w} \leq \vec{b}$$

For simple problems, we can sometimes describe the feasible points geometrically.

Suppose that we want to find the 2D vector that is closest to the point $(3, 2)$ and that satisfies three constraints:

$$\begin{aligned} w_1 \geq 0 \qquad & \leftrightarrow \qquad \begin{bmatrix} -1 & 0 \end{bmatrix} \vec{w} \leq 0 \\ w_2 \geq 0 \qquad & \leftrightarrow \qquad \begin{bmatrix} 1 & 0 \end{bmatrix} \vec{w} \leq 3 \\ w_1 + w_2 \leq 4 \qquad & \leftrightarrow \qquad \begin{bmatrix} 1 & 1 \end{bmatrix} \vec{w} \leq 4 \end{aligned}$$

The individual constraints can be visualized as half-planes. The intersection of these convex half-planes is a convex set which, in this case, describes a triangle in the plane. In Figure 24.3(A-C), the half of the plane that satisfies each individual constraint is shown as being shaded and unshaded regions indicate points that violate each constraint. The intersection of the constraints is shown in Figure 24.3(D).

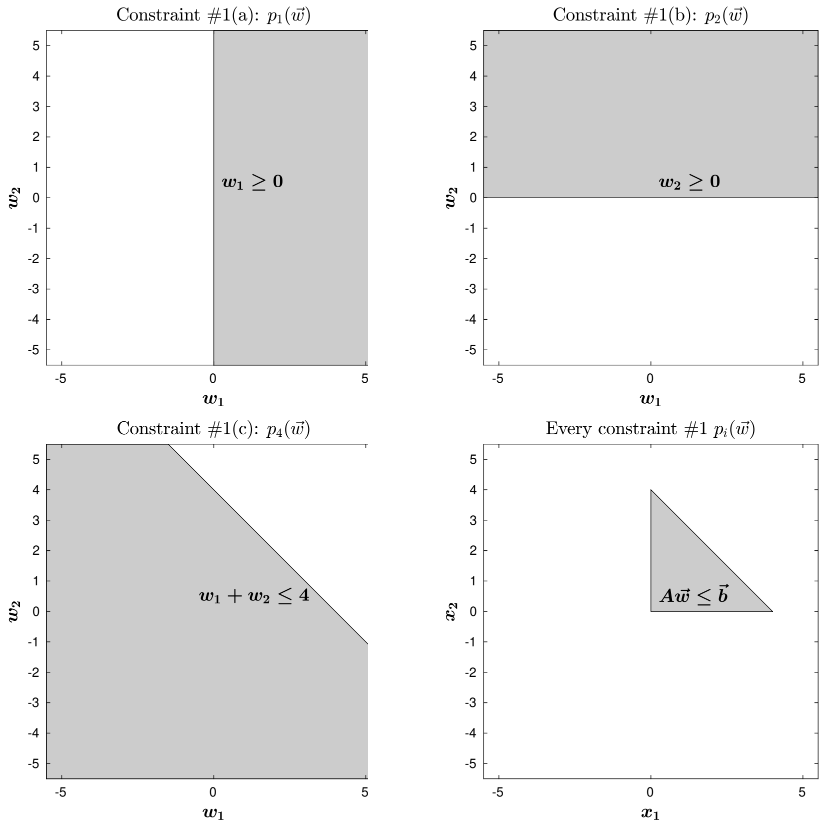


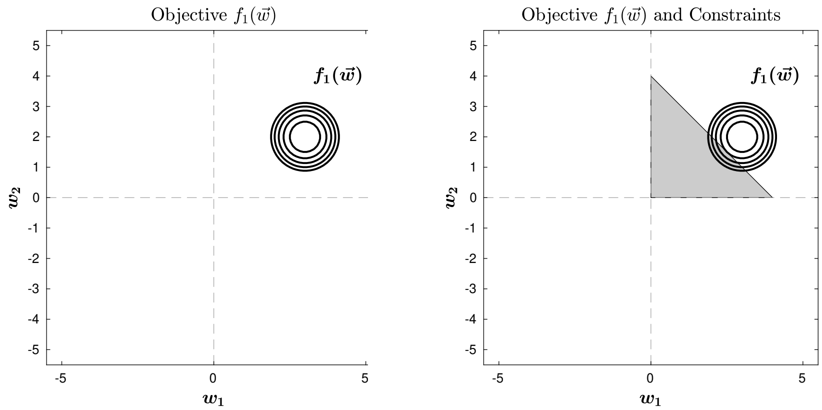
Figure 24.3: Three linear inequality constraints, individually and jointly satisfied. The shaded region depicts the feasible points for each constraint. (A) $w_1 \geq 0$. (B) $w_2 \geq 0$. (C) $w_1 + w_2 \leq 4$. (D) The shaded region shows the feasible points.

Contours of the objective function, alone and with respect to the constraints, are shown in Figure 24.4. A solution to the problem is a feasible point with a minimum distance to the desired point.

24.2 Specific Solutions and General Solutions

Determining the feasible points for a problem may be challenging. In current textbooks and notes for graduate-level courses in optimization, there are five common methods for solving a quadratic objective with linear inequality constraints:

Specific Solution: The objective and constraints are analyzed, for a specific problem, to find a specific solution. This method is used for problems where the results of the analysis lead to a computationally effective implementation.



(A) **(B)**

Figure 24.4: An objective function and four linear inequality constraints. (A) Contours of the objective, which is to minimize the Euclidean distance to a desired point. (B) Contours of the objective and the set of feasible points, which are in the shaded region.

Active Sets: From an initial feasible solution, the feasible set is explored to determine the constraints that are *active*, which are constraints for which an equality relation holds. The equality constraints can be managed using the augmented matrix method that was presented earlier in the course.

Interior Points: Each inequality constraint is replaced with a *barrier function* that prevents a descent algorithm from violating the constraints. This requires an initial estimate of the minimizer that is a feasible point, and then converts the constrained minimization problem into an unconstrained problem.

Penalty Methods: These use a different form of the Lagrange function so that the constraint term is convex. The Lagrange multipliers $\vec{\lambda}$ can then be treated as free variables, which converts the constrained minimization problem into an unconstrained problem. These also require an initial estimate of the minimizer that is a feasible point.

Sequential Quadratic Programming: The augmented matrix for an equality-constrained problem is solved by numerical approximation. Each iteration is a quadratic problem where the possibly dense matrix K is approximated with a symmetric positive indefinite matrix that is easier to solve.

We will use specific solutions to specific problems in this course. Our solutions, and many of the other solutions, rely on necessary conditions for a minimizer that are also sufficient conditions for linear inequality constraints.

Inequality constraints, for nonlinear functions as well as for linear inequalities, were characterized by Harold Kuhn and Albert Tucker (Princeton University and Stanford University, respectively) in 1951. The necessary and sufficient conditions for constrained optimization of a locally convex function are sometimes called the Kuhn-Tucker conditions. Because the necessary conditions were – unknown to Kuhn and Tucker – published in a Master's thesis by William Karush in 1931 (University of Chicago), these are widely known as the KKT conditions.

24.3 KKT Conditions for a Quadratic Objective and Linear Inequalities

A KKT point for Problem 24.2 is a feasible point that: has non-negative Lagrange multipliers λ_i; satisfies the Lagrange equation for the problem; and satisfies a condition in the Lagrange multipliers that is often called complementary slackness.

Definition: KKT point

For a primal problem of a quadratic objective with linear inequality constraints, having the form of Problem 24.2, that a vector $\hat{w}$ with associated Lagrange multipliers $\hat{\lambda}$ is a *KKT point* is defined as a point that satisfies the four conditions

$$\begin{array}{lrcl} \text{(1) Primal feasibility:} & A\hat{w} & \leq & \vec{b} \\ \text{(2) Dual feasibility:} & \exists\hat{\lambda} \in \mathbb{R}^m & \geq & \vec{0} \\ \text{(3) Stationarity:} & K\hat{w} + \vec{q} + A^T\hat{\lambda} & = & \vec{0} \\ \text{(4) Complementary slackness:} & \hat{\lambda}^T\left[A\hat{w} - \vec{b}\right] & = & \vec{0} \end{array} \tag{24.9}$$

Kuhn and Tucker proved a necessary condition that applies to any continuous differentiable $f : \mathbb{R}^n \to \mathbb{R}$ and any continuous differentiable $p : \mathbb{R}^n \to \mathbb{R}^m$. They showed that if $\vec{w}^*$ is a local minimizer of f, then $\vec{w}^*$ is a KKT point.

A sufficient condition is that, at a KKT point $\hat{w}$, the objective function is convex. This is often stated as

$$\nabla^2 f(\hat{w}) \succeq 0$$

For Problem 24.2, we have a specific version of the general result. Because the objective function f is quadratic in $\vec{w}$, it is convex; and because the inequality $\vec{p}$ is an affine inequality, it is convex. If such a problem is consistent – that is, if the feasible set is not the empty set – then there is a solution to the constrained problem.

Theorem: KKT condition for Problem 24.2

$$\textit{For Problem 24.2, } \vec{w}^* \textit{ is a global minimizer if and only if } \vec{w}^* \textit{ is a KKT point} \tag{24.10}$$

Proof: See Beck, 2014, p. 196

Observation: This theorem allows us to focus on the dual problem, which in many applications is computationally preferred.

We will use specific solutions to inequality constraints in some examples of optimization problems, such as constrained least-squares problems and "denoising" of data. Later in this course, we will explore how the KKT conditions and the dual formulation can be used to find an optimal hyperplane that separates two sets of labeled points.

References

[1] Beck A: Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. Siam Press, 2014

[2] Chong EKP, Zak SH: An Introduction to Optimization, volume 76. John Wiley & Sons, 2013

[3] Nocedal J, Wright S: Numerical Optimization. Springer Science & Business Media, 2006