CISC 371 Class 20

Lagrange Multipliers for 2D Functional Convex Problems

Texts: [1] pp. 285–296; [2] pp. 195–203; [3] p. 215; [4] pp. 374–391; [5] pp. 304–315

Main Concepts:

- *A minimizer is in a level curve of an objective*
- *A minimizer is in a level curve of a property*
- *A minimizer is in the intersection of level curves*
- *At the intersection point, example level curves have parallel gradients*

Sample Problem, Machine Inference: What is the minimum-length vector on a line in 2D?

In a computational setting, the constraints on an optimization problem are usually given as functions. We will call each constraint function a *property*, writing these as:

$$\begin{array}{rrcl} \textit{Equality Property:} & p(\vec{w}) & = & 0 \\ \textit{Inequality Property:} & p(\vec{w}) & \leq & 0 \end{array} \tag{20.1}$$

We can see that a property is closely related to a previous concept: an equality property is a level curve and an inequality property is a level set. A functional constrained optimization problem has a minimizer – if one exists – that is in a level curve or a level set.

How can we minimize an objective, subject to constraints? We can begin by exploring the geometry of some simple problems for size-2 vectors. We can start with the problem of finding a vector with minimum squared length that satisfies a linear equality property.

<u>Problem:</u> squared length, linear equality

$$\begin{array}{lrcl} & \vec{w}^* & = & \operatorname*{argmin}_{\vec{w}\in\mathbb{R}^2} f_1(\vec{w}) \\ & p_1(\vec{w}^*) & = & 0 \\ \text{where:} & & & \\ & f_1(\vec{w}) & = & \vec{w}^T\vec{w} \\ & p_1(\vec{w}) = \begin{bmatrix} -1 & 1 \end{bmatrix}\vec{w} + 3 & = & 0 \end{array} \tag{20.2}$$

Recall that the level curves of a function with a size-2 argument are curves in the plane, so they can be superimposed as a contour plot. We can plot the property $p_1(\vec{w})$ and the level curves of f_1 to get a picture of Problem 20.2. This combined plot is shown in Figure 20.1.

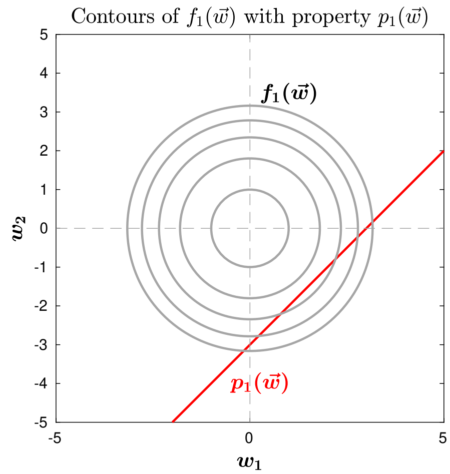


Figure 20.1: Plots of property p_1, shown in black, and level curves of function f_1, shown in blue. The minimizer of f_1 is on a level curve that intersects the curve of property p_1.

The optimal vector $\vec{w}^*$ for Problem 20.2 must satisfy property p_1, so it must be on the level curve $p_1(\vec{w}) = 0$; we can write this as

$$\vec{w}^* \in \mathbb{S}_C(p_1, 0)$$

When we evaluate f_1 using the optimal vector, we get some scalar value $c_1 = f_1(\vec{w}^*)$. This has a useful implication:

$\vec{w}^*$ *is on the level curve of* f_1 *for value* c_1, or $\vec{w}^* \in \mathbb{S}_C(f_1, c_1)$

The optimal vector $\vec{w}^*$ for Problem 20.2 must be on the level curve of p_1 at value 0, and on the level curve of f_1 at value c_1. This intersection is illustrated in Figure 20.2(A). We can see, in Figure 20.2 (A), that the level curves "touch": they are not crossing. If we plot the transpose of the gradients at the optimal point, that is if we plot $[\underline{\nabla} f_1(\vec{w}^*]^T$ and $[\underline{\nabla} p_1(\vec{w}^*]^T$, we can see that the transposed gradients are parallel and are in opposite directions. This is illustrated in Figure 20.2(B).

We can compute the gradients of f_1 and p_1 as

$$\begin{aligned} \underline{\nabla} f_1(\vec{w}) &= \begin{bmatrix} 2w_1 & 2w_2 \end{bmatrix} \\ \underline{\nabla} p_1(\vec{w}) &= \begin{bmatrix} -1 & 1 \end{bmatrix} \end{aligned} \tag{20.3}$$

The optimal point is

$$\vec{w}_1^* = \begin{bmatrix} 1.5 \\ -1.5 \end{bmatrix} \tag{20.4}$$

At the optimal point $\vec{w}^*$, the gradients are a scalar multiple $\mu_1 = 3$ of each other, so

$$\underline{\nabla} f_1(\vec{w}_1^*) = \mu_1 \underline{\nabla} p_1(\vec{w}_1^*) \tag{20.5}$$

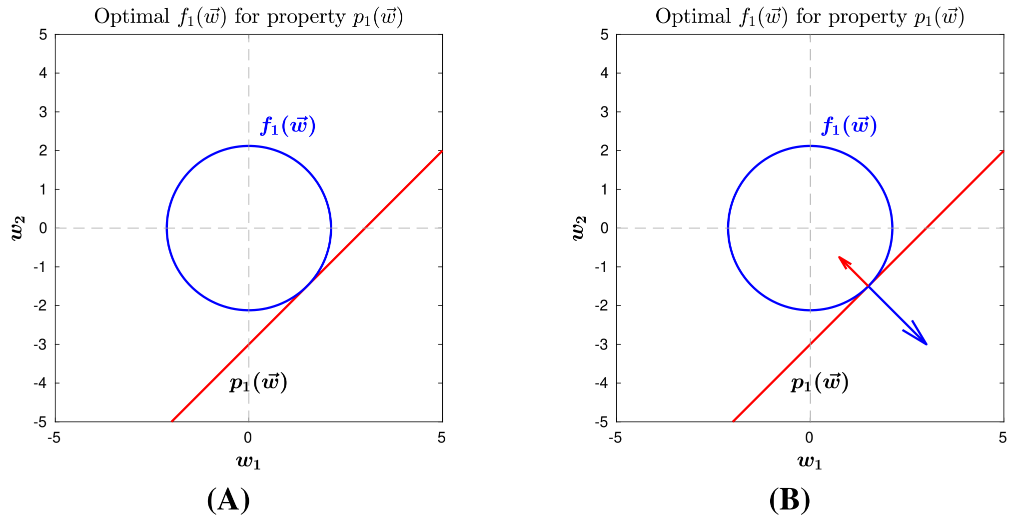


Figure 20.2: Plots of property p_1, shown in black, and the optimal level curve of function f_1, shown in green. (A) The minimizer $\vec{w}^*$ is on a level curve of objective f_1 that intersects the level curve of property p_1. (B) The respective transposed gradient, shown as arrows, are parallel and opposite in direction.

<u>**Problem:**</u> weighted squared length, quadratic equality

$$
\begin{aligned}
\vec{w}^* &= \operatorname*{argmin}_{\vec{w}\in\mathbb{R}^2} f_2(\vec{w}) \\
p_2(\vec{w}^*) &= 0 \\
\text{using:} \qquad R &= \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \\
K &= R \begin{bmatrix} 1/(3^2) & 0 \\ 0 & 1/(2^2) \end{bmatrix} R^T \\
\text{where:} \qquad & \\
f_2(\vec{w}) &= \frac{1}{2}\vec{w}^T K \vec{w} \\
p_2(\vec{w}) = (w_1 - 3)^2 + 1 - w_2 &= 0
\end{aligned}
\tag{20.6}
$$

This problem is to minimize a weighted norm, subject to a quadratic equality constraint. As before, we can plot the level curve of property p_2 at value 0, which is $\mathbb{S}_C(p_2, 0)$. We can also plot the contours of f_2, which are level curves at various values. This is illustrated in Figure 20.3.

As with Problem 20.2, the optimal point $\vec{w}^*$ to Problem 20.6 must be in the level curve of property p_2 at value 0, so $\vec{w}^* \in \mathbb{S}_C(p_2, 0)$. The optimal point must be in a level curve of f_2 at some value c_2, so $\vec{w}^* \in \mathbb{S}_C(f_2, c_2)$. This geometry is illustrated in Figure 20.4(A) where, once again, the level curves "touch". The transposed gradients are parallel and are in opposite directions, as illustrated in Figure 20.4(B).

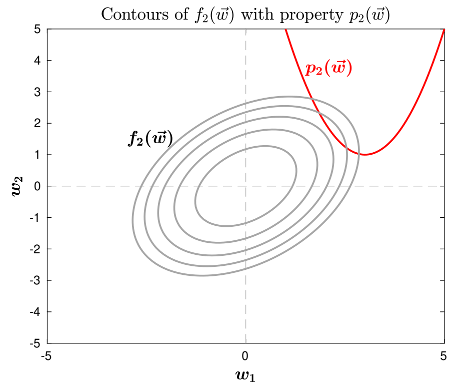
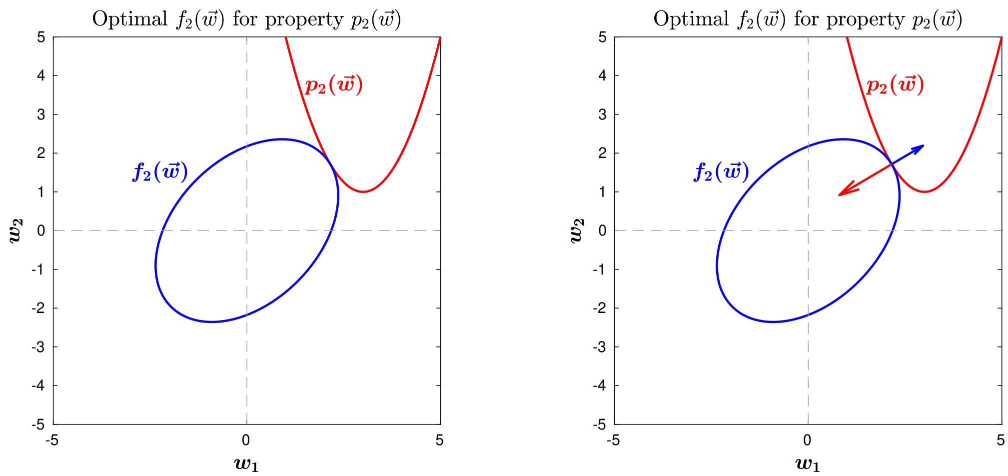


Figure 20.3: Plots of property p_2, shown in black, and level curves of function f_2, shown in blue. The minimizer of f_2 is on a level curve that intersects the curve of property p_2.



(A) **(B)**

Figure 20.4: Plots of property p_2, shown in black, and the optimal level curve of function f_2, shown in green. (A) The minimizer $\vec{w}^*$ is on a level curve of objective f_2 that intersects the level curve of property p_2. (B) The respective transposed gradients, shown as arrows, are parallel and opposite in direction.

We can compute the gradients of f_2 and p_2 as

$$\begin{aligned} \underline{\nabla} f_2(\vec{w}) &= [\vec{w}]^T K \\ \underline{\nabla} p_2(\vec{w}) &= \begin{bmatrix} 2w_1 - 6 & -1 \end{bmatrix} \end{aligned} \tag{20.7}$$

The optimal point is

$$\vec{w}_2^* \approx \begin{bmatrix} 2.155 \\ 1.714 \end{bmatrix} \tag{20.8}$$

At the optimal point $\vec{w}_2^*$, the gradients are a scalar multiple $\mu_2 \approx -6.257$ of each other, so

$$\underline{\nabla} f_2(\vec{w}_2^*) = \mu_2 \underline{\nabla} p_2(\vec{w}_2^*) \tag{20.9}$$

<u>Problem:</u> affine objective, quadratic constraint

$$\begin{aligned} \vec{w}^* &= \operatorname*{argmin}_{\vec{w}\in\mathbb{R}^2} f_3(\vec{w}) \\ p_3(\vec{w}^*) &= 0 \\ \text{where:} & \\ f_3(\vec{w}) &= \begin{bmatrix}1 & 1\end{bmatrix} \vec{w} \\ p_3(\vec{w}) = (w_1-1)^2 + 1 - w_2 &= 0 \end{aligned} \tag{20.10}$$

This problem is to minimize an affine form, subject to a quadratic equality constraint. As before, we can plot the level curve of property p_3 at value 0, which is $\mathbb{S}_C(p_3, 0)$. We can also plot the contours of f_3, which are level curves at various values. This is illustrated in Figure 20.5.

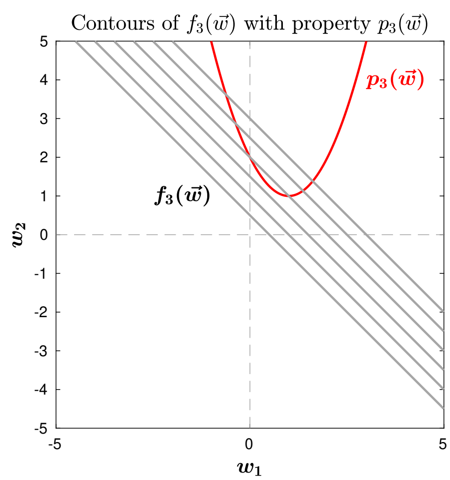
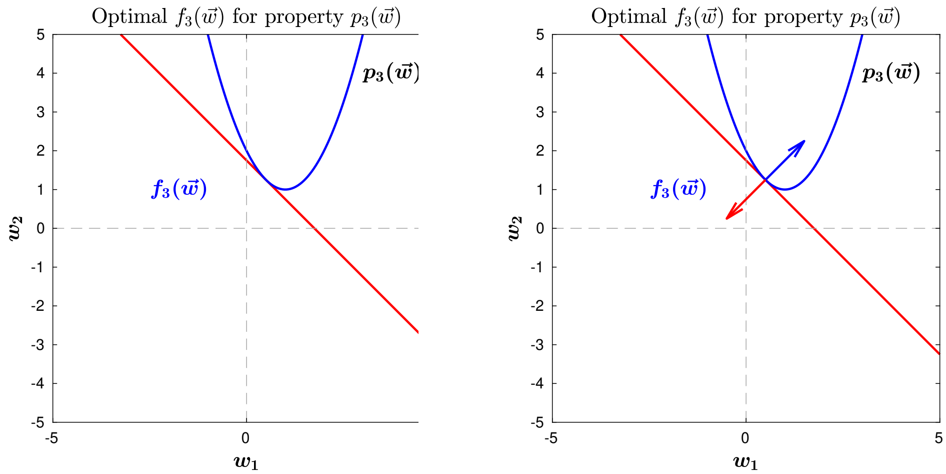


Figure 20.5: Plots of property p_3 are shown in black; level curves of function f_3, which are lines, are shown in blue. The minimizer of f_3 is on a level curve that intersects the curve of property p_3.

As with Problem 20.2, the optimal point $\vec{w}^*$ to Problem 20.10 must be in the level curve of property p_3 at value 0, so $\vec{w}^* \in \mathbb{S}_C(p_3, 0)$. The optimal point must be in a level curve of f_3 at some value c_3, so $\vec{w}^* \in \mathbb{S}_C(f_3, c_3)$. This geometry is illustrated in Figure 20.6(A) where, once again, the level curves "touch"; the transposed gradients are parallel and are in opposite directions. This is illustrated in Figure 20.6(B).



(A) **(B)**

Figure 20.6: Plots of property p_3, shown in black, and the optimal level curve of function f_3, shown in green. (A) The minimizer $\vec{w}^*$ is on a level curve of objective f_3 that intersects the level curve of property p_3. (B) The respective transposed gradients, shown as arrows, are parallel and opposite in direction.

We can compute the gradients of f_3 and p_3 as

$$\begin{aligned} \underline{\nabla} f_3(\vec{w}_3^*) &= \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \underline{\nabla} p_3(\vec{w}_3^*) &= \begin{bmatrix} 2w_1 - 2 & -1 \end{bmatrix} \end{aligned} \tag{20.11}$$

The optimal point is

$$\vec{w}_3^* = \begin{bmatrix} 0.5 \\ 1.25 \end{bmatrix} \tag{20.12}$$

At the optimal point $\vec{w}^*$, the gradients are a scalar multiple $\mu_3 = -1$ of each other, so

$$\underline{\nabla} f_3(\vec{w}_3^*) = \mu_3 \underline{\nabla} p_3(\vec{w}_3^*) \tag{20.13}$$

20.1 Lagrange Multipliers

In exploring these examples, we have observed that:

- The minimizer $\vec{w}^*$ was in a level curve of the objective f
- The minimizer $\vec{w}^*$ was in a level curve of the property p
- The gradients of f and p at $\vec{w}^*$ were related:

$$\underline{\nabla} f(\vec{w}^*) = -\mu \underline{\nabla} p(\vec{w}^*)$$

These results extend from size-2 vectors, which are in the Euclidean space $\mathbb{R}^2$, to size-n vectors, which are in the Euclidean space $\mathbb{R}^n$. For reasons having to do with the existence of the directional derivative at interior points, the observations regarding level curves are extended to level *sets*.

The scalar $\mu \in \mathbb{R}$ is called the *Lagrange multiplier*. We only need a version of the multiplier theorem that applies to any vector $\vec{w} \in \mathbb{R}^n$; there are versions that apply to subsets of vector spaces, and to non-Euclidean spaces such as manifolds.

Theorem: Lagrange multiplier μ

For any continuous differentiable $f : \mathbb{R}^n \to \mathbb{R}$, and any continuous differentiable $p : \mathbb{R}^n \to \mathbb{R}$, and any $\vec{w}_0 \in \mathbb{R}^n$, where the level set of p at $\vec{w}_0$ is $\mathbb{S}_S(p, p(\vec{w}_0))$,

If $\vec{w}_0$ is a local minimizer or a local maximizer of f on the domain $\mathbb{S}_S(p, p(\vec{w}_0))$,

then there is a scalar $\mu \in \mathbb{R}$ such that

$$\underline{\nabla} f(\vec{w}_0) = -\mu \underline{\nabla} p(\vec{w}_0) \tag{20.14}$$

Proof: Most textbooks on vector calculus

Observation: The proof of Theorem 20.14 typically involves the definition of the tangent space of a level set, and the Implicit Function Theorem in a vector space.

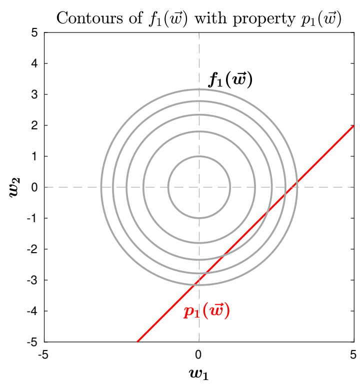
Extra Notes

Extra Notes on 2D Convex Problems

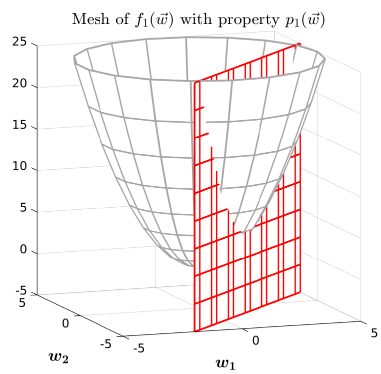
Earlier in the course, we observed that a 2D function can be usefully visualized using contour plots and a surface plot. For example, Figure 20.1 provides contour plots of the objective function

f_1 in blue and the curve of the property function p_1 in blue. What would this look like as a surface rendering? Plotting the objective function is straightforward; it may not be immediately clear how to plot a 2D constraint curve in 3D.

We stated the constraint as property $p_1(\vec{w})$ that evaluates to zero. This is very close to what we mean a level curve curve to be: a point $\vec{w}$ is in the level curve if and only if the property function evaluates to zero. Because the 2D property does or does not hold, independent of the value of $f_1(\vec{w})$, we can render the level curve by "extending" the level curve to every value in $\mathbb{R}$ that the objective function could possibly map to. Because $p_1(\vec{w}) = 0$ is a line, the infinite extension of the line is a vertical plane. This is shown in Figure 20.7.



(A)

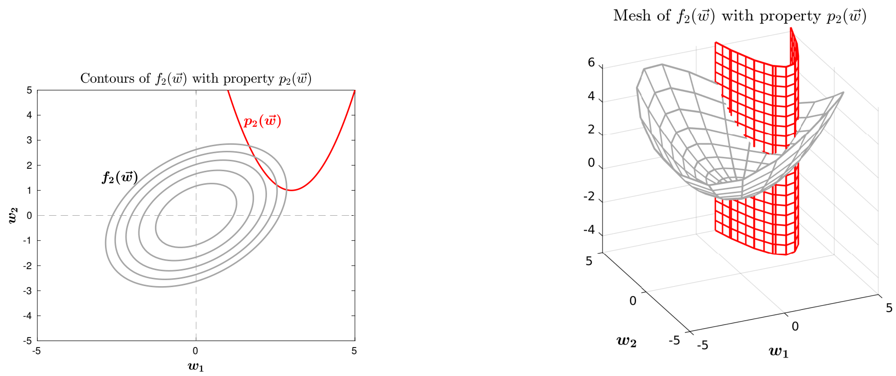


(B)

Figure 20.7: Plots of an objective function f_1, shown in blue, and a constraint property function p_1, shown in black. The minimizer of f_1 is on a level curve that intersects the curve of property p_1. (A) Contour plots, which are level curves of the objective function in blue and the level curve of the property function at zero. (B) Surface plots of the objective function in blue and the infinitely extended constraint property.

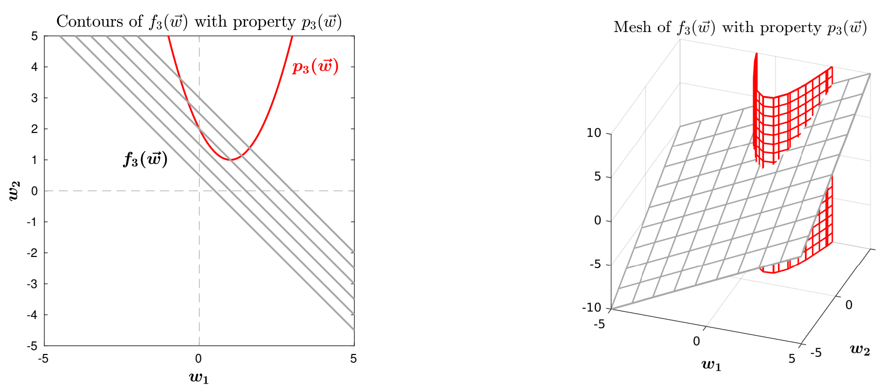
We can use the same concept to visualize a quadratic objective f_2 function and a parabolic constraint function p_2. The constraint function will now be rendered as a vertical parabolic sheet, shown in Figure 20.8.

When we visualize an affine objective f_3 function and a parabolic constraint function p_3, we can see how a constrained optimization problem can have a solution even when the unconstrained objective function tends to $-\infty$. The constraint function will be rendered as a vertical parabolic sheet and the objective function will be rendered as a "tilted" plane, shown in Figure 20.9. Although the plane tends to $-\infty$, the points in the plane that satisfy the parabolic constraint have a unique solution.



(A) **(B)**

Figure 20.8: Plots of an objective function f_2, shown in blue, and a constraint property function p_2, shown in black. The minimizer of f_2 is on a level curve that intersects the curve of property p_2. (A) Contour plots, which are level curves of the objective function in blue and the level curve of the property function at zero. (B) Surface plots of the objective function in blue and the infinitely extended constraint property.



(A) **(B)**

Figure 20.9: Plots of an objective function f_2, shown in blue, and a constraint property function p_2, shown in black. The minimizer of f_2 is on a level curve that intersects the curve of property p_2. (A) Contour plots, which are level curves of the objective function in blue and the level curve of the property function at zero. (B) Surface plots of the objective function in blue and the infinitely extended constraint property.

References

[1] Antoniou A, Lu WS: Practical Optimization: Algorithms and Engineering Applications. Springer Science & Business Media, 2007

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

[3] Boyd S, Vandenberghe L: Convex Optimization. Cambridge University Press, 2004

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

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