# Lecture 17: MAP Estimation with Non-Gaussian Priors

Chapter 7

Charles A. Bouman

Purdue University

Friday, October 2, 2026

# Outline

1. The MAP cost function and the positivity constraint.
2. Convexity: when the minimum exists, is unique, and is stable.
3. Line search, and the three algorithms built from it.
4. ICD, and how it enforces positivity.
5. Solving the scalar line search when the prior is non-quadratic.

Reading: Chapter 7, *Foundations of Computational Imaging*.

# Today's Big Ideas

- A non-Gaussian prior makes the MAP cost nonquadratic.
- Convexity makes the minimum unique and stable.
- Three methods: one line search, different directions
- Along a line, only the prior is nonquadratic. The exact step comes from a slow search, which the next chapter fixes.

# Why we need new optimization methods

A non-Gaussian prior makes the MAP cost function nonquadratic, so we need **new methods for non-quadratic optimization**.

- **Convexity** tells us that the MAP problem has a unique and computable solution.
- An efficient **line search** matters: these problems are large
- Often the image must be nonnegative: the **positivity constraint**

$$\Omega = \mathbb{R}^{+N} = \left\{ x \in \mathbb{R}^N : x_i \geq 0 \text{ for all } i \right\}$$

# The MAP cost function, and the three steps

The MAP estimate of $X$ given $Y$ is

$$\hat{x} = \arg\min_{x \in \Omega} f(x; y) \ , \qquad f(x; y) = -\log p_{y|x}(y|x) - \log p(x) \ ,$$

where $f(x; y)$ is the **MAP cost function**.

So MAP estimation has three steps:

1. Select the forward model, $p_{y|x}(y|x)$.
2. Select the prior model, $p(x)$.
3. Minimize the cost function.

## A caveat about step 3

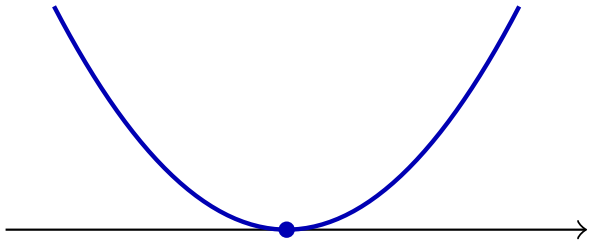
In principle only steps 1 and 2 set the quality of the estimate. In practice the optimizer does not reach the exact minimum, so step 3 affects the quality too.

# Convexity: uniqueness and stability of the MAP estimate

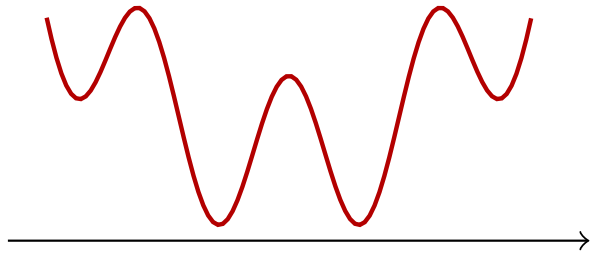
Convexity of $f(x; y)$ in $x$ decides whether the MAP estimate is unique and stable.

## The key results (Appendix A)

For a *convex* function, any local minimum is also a global minimum. For a *strictly convex* function, the global minimum is unique, and the MAP estimate is a continuous (stable) function of the data $y$.



convex



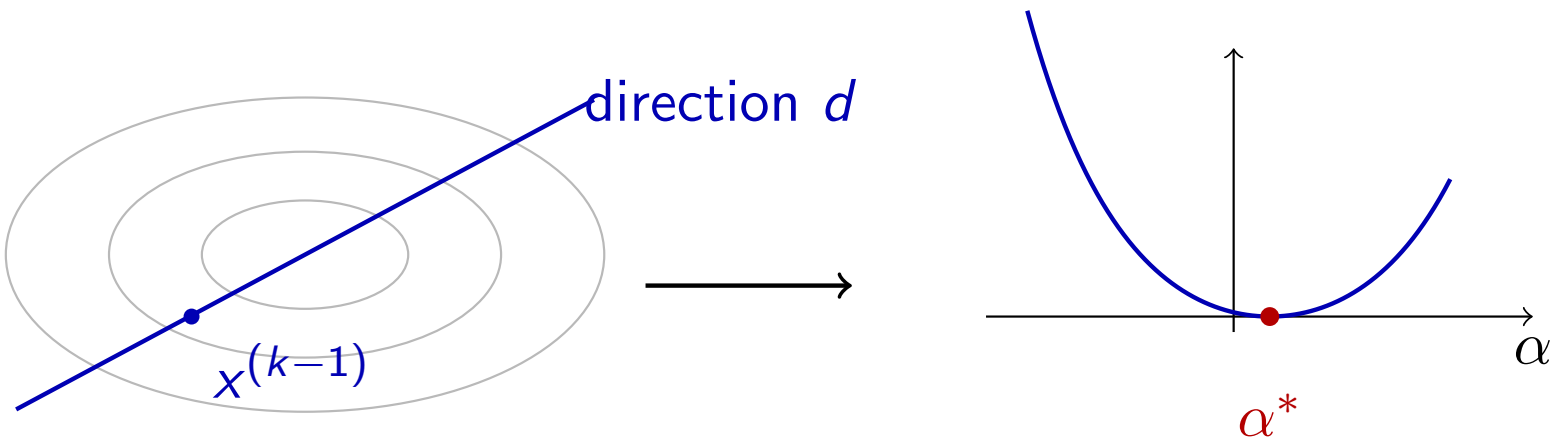
nonconvex

This matters because most practical optimizers reach, at best, a *local* minimum.

# Line search: reduce to one dimension

A **line search** fixes a direction $d \in \mathbb{R}^N$ and minimizes the cost along the line $x^{(k-1)} + \alpha d$ over the single scalar step size $\alpha \in \mathbb{R}$:

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}} f\left(x^{(k-1)} + \alpha d\right) , \qquad x^{(k)} = x^{(k-1)} + \alpha^* d \ .$$




$N$ dimensions

one dimension

For a general prior the one-dimensional cost is not quadratic, so $\alpha^*$ has no closed form. We must search for it.

# One framework, three algorithms

The three algorithms differ only in how they choose the direction $d$.

- **Steepest descent** (Lecture 13): negative gradient, $d = -\nabla f(x)$
- **Conjugate gradient** (Lecture 14): the negative gradient is adjusted to be conjugate to all previous directions.
- **ICD** (Lecture 14): one pixel per update, $d = \varepsilon_s$

## Why efficient line search matters

The line search is common to all three. For nonquadratic problems it is the main computational burden, so an efficient line search is very important.

# ICD, also known as Gauss–Seidel iteration

A full ICD iteration updates each pixel once, in a fixed order:

$$\begin{aligned} &\text{For } s \in S \quad \{ \\ &\qquad \alpha^* \leftarrow \arg\min_{\alpha \in \mathbb{R}} f(x + \alpha \varepsilon_s) \\ &\qquad x \leftarrow x + \alpha^* \varepsilon_s \\ &\qquad \} \end{aligned}$$

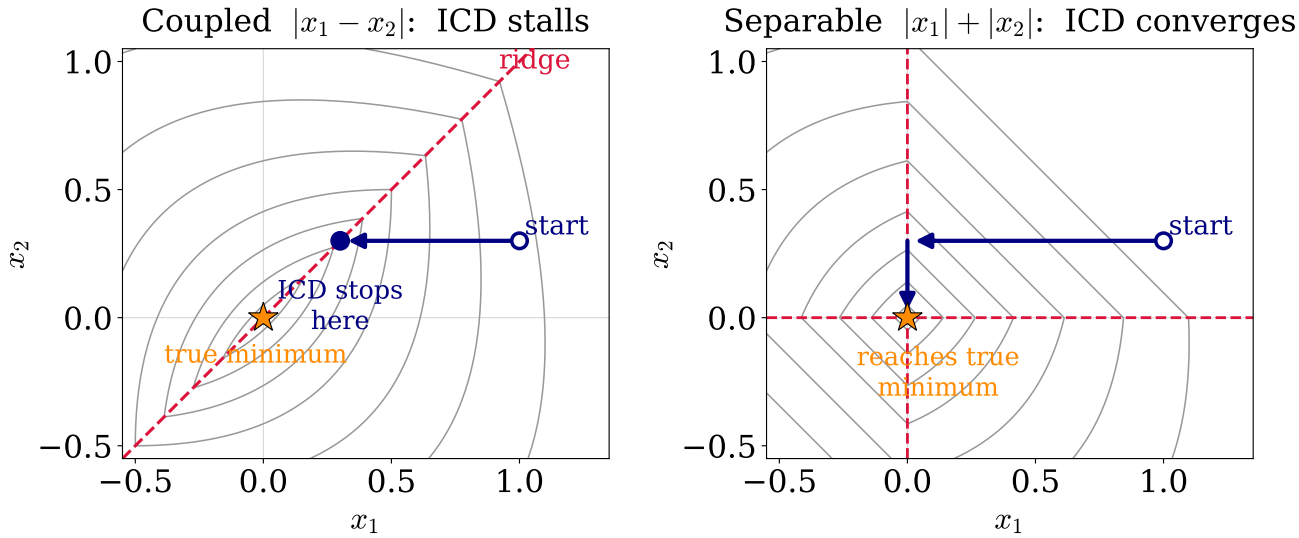
where $\varepsilon_s$ is the vector that is one in component $s$ and zero elsewhere, so each step changes only pixel $x_s$.

**Theorem: ICD Convergence**

Let $f : \Omega \to \mathbb{R}$ be strictly convex and continuously differentiable on the convex set $\Omega \subset \mathbb{R}^N$. Then the ICD iterations converge to the global minimum of $f$.

# ICD can stall when the cost is not differentiable

The convergence theorem assumes a differentiable cost. A convex but non-differentiable cost can make ICD stall before it reaches the minimum.




- Left: $|x_1 - x_2|$ couples the pixels, so ICD stalls on the ridge
- Right: $|x_1| + |x_2|$ is separable, so ICD converges
- ICD converges for any $f(x) = \frac{1}{2} x^t B x + b^t x + \|x\|_1$ with $B$ positive definite. [Alliney and Ruzinsky, 1994]

# Enforcing positivity with ICD

ICD enforces positivity by restricting each pixel to be nonnegative.

**ICD with positivity:**

For $K$ iterations $\{$

  For each pixel $s \in S$ $\{$

    $\alpha^* \leftarrow \arg\min_{\alpha \geq -x_s} f(x + \alpha \varepsilon_s)$

    $x_s \leftarrow x_s + \alpha^*$

  $\}$

$\}$

- Other methods also enforce constraints:
  - KKT: general conditions for constrained optimization.
  - ADMM (Chapter 9): current state-of-the-art.

# Line search with a non-quadratic prior

In order to do a line search, we must solve

$$\alpha^* = \arg\min_{\alpha} f(x + \alpha d) \ .$$

But $f(x + \alpha d)$ is not quadratic, so we split it into a quadratic data term and a non-quadratic prior term:

$$f(x + \alpha d) = c + \underbrace{\theta_1 \alpha + \tfrac{1}{2}\theta_2 \alpha^2}_{\text{data term (quadratic)}} + \underbrace{\sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \rho(\Delta x_{s,r} + \alpha \Delta d_{s,r})}_{\text{prior (non-quadratic)}} \ , \qquad (1)$$

where $\Delta x_{s,r} = x_s - x_r$, $\Delta d_{s,r} = d_s - d_r$, and

$$\theta_1 = -e^t \Lambda A d \ , \qquad \theta_2 = d^t A^t \Lambda A d \ , \qquad e = y - Ax \ .$$

Solution is given by:

$$\alpha^* = \arg\min_{\alpha} \left\{ \theta_1 \alpha + \tfrac{1}{2}\theta_2 \alpha^2 + \sum_{\{s,r\} \in \mathcal{P}} b_{s,r} \, \rho(\Delta x_{s,r} + \alpha \Delta d_{s,r}) \right\} \ .$$

# Example 7.1: ICD with a nonquadratic prior

For ICD the direction selects one pixel, $d = \varepsilon_s$, so $\Delta d_{s,r} = 1$ for every neighbor $r$ of $s$, and

$$\theta_1 = -e^t \Lambda A_{*,s} \ , \qquad \theta_2 = A_{*,s}^t \Lambda A_{*,s} \ ,$$

where $A_{*,s}$ is the $s^{th}$ column of $A$. Only pixel $x_s$ changes, so only its neighbors enter the prior sum:

$$\alpha^* = \underset{\alpha \geq -x_s}{\arg\min} \left\{ \theta_1 \alpha + \frac{1}{2} \theta_2 \alpha^2 + \sum_{r \in \partial s} b_{s,r} \rho(x_s - x_r + \alpha) \right\} \ , \tag{2}$$

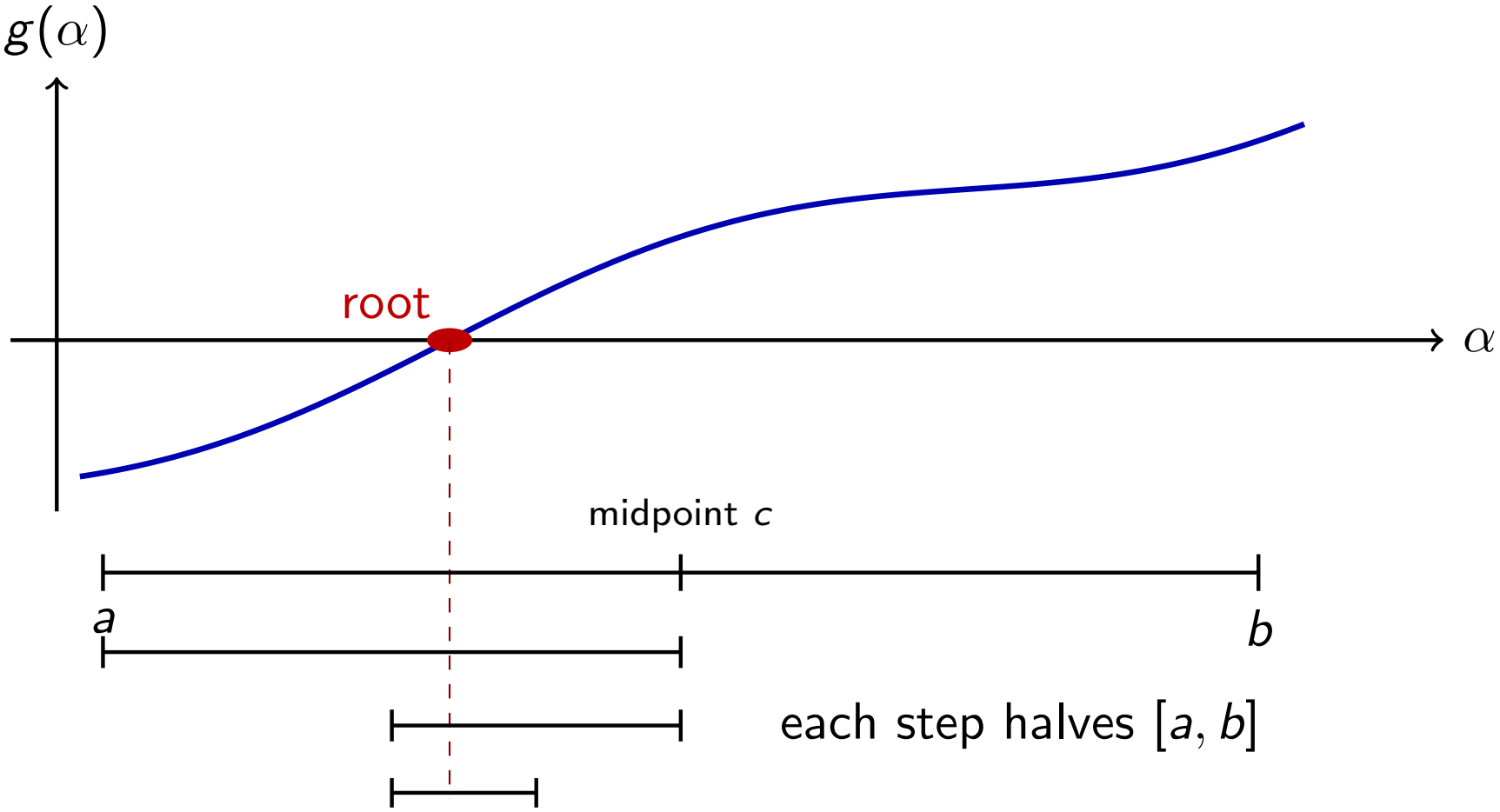
where $\partial s$ is the set of neighbors of $s$, and $\alpha \geq -x_s$ keeps the pixel nonnegative.

The pixel and the error are then updated:

$$x_s \leftarrow x_s + \alpha^* \ , \qquad e \leftarrow e - A_{*,s} \alpha^* \ .$$

# Finding the minimum by rooting the derivative

The minimum is where the derivative $g(\alpha) \triangleq \frac{\partial}{\partial \alpha} f(x + \alpha d)$ is zero. When the potentials are strictly convex, $g$ is strictly increasing, so $g(a) < 0$ and $g(b) > 0$ bracket a single root. Halving the bracket closes in on it.



# The half-interval search algorithm

Each iteration halves the bracket $[a, b]$ by testing the sign of $g$ at the midpoint $c = (a + b)/2$.

**Half-Interval Line-Search Algorithm:**

Initialize $a$ and $b$ so that $g(a) < 0$ and $g(b) > 0$.

For $k = 1$ to $K$ $\{$

$c \leftarrow (a + b)/2$

If $g(c) \leq 0$, then $a \leftarrow c$

If $g(c) > 0$, then $b \leftarrow c$

$\}$

$$\alpha^* = \frac{-g(a)}{g(b) - g(a)}(b - a) + a$$

**Do not skip the final interpolation step (the last line)**

The endpoints land on a grid of width $(b - a)/2^K$. The interpolation places $\alpha^*$ between them, close to the true root.

# Two disadvantages of half-interval search

## 1. The bracket must be known in advance

The method needs a finite interval $[a, b]$ with $g(a) < 0$ and $g(b) > 0$. If the root is not bracketed, one must first widen the interval, and each check costs another evaluation of $g$.

## 2. Evaluating $g(\alpha)$ is expensive

Each evaluation of $g(\alpha)$ takes significant computation. Repeated evaluations may be an unacceptable overhead. It helps to keep the cost function **monotonically decreasing**, so that stopping early still lowers the cost.

# Summary

- A non-Gaussian prior makes the MAP cost nonquadratic.
- Convexity: unique minimum, and an estimate stable in the data
- Steepest descent, conjugate gradient, and ICD are one line search with directions $-\nabla f$, a conjugate direction, or $\varepsilon_s$.
- Along the line, the data term is quadratic; the prior is not
- The exact step roots $g(\alpha) = 0$ by half-interval search, which is slow.

**Reading:** Chapter 7. **Next lecture:** Chapter 8, surrogate functions.