Proximal Methods in Numerical Optimization

Lecture II – Proximal Point Methods

Alberto DE MARCHI
University of the Bundeswehr Munich
alberto.demarchi@unibw.de
aldma.github.io

Povo, UniTN – March 3, 2026

⚠ these slides are under development: please email me for corrections and suggestions ⚠

Outline

Setting

Problem

Let $f : \mathcal{H} \to (-\infty, +\infty]$ be a **proper, lower semicontinuous** function on a Hilbert space $\mathcal{H}$. We want to solve

$$\operatorname*{minimize}_{x \in \mathcal{H}} \quad f(x).$$

The **proximal operator** of f with parameter $\gamma > 0$ is

$$\operatorname{prox}_{\gamma f}(x) := \operatorname*{arg\,min}_{u \in \mathcal{H}} \left\{ f(u) + \frac{1}{2\gamma} \|u - x\|^2 \right\}.$$

For f **uniformly prox-bounded**, $\operatorname{prox}_{\gamma f}$ is well-defined for all $\gamma \in (0, \gamma_f)$.

For f **convex**, it is also *single-valued* and *nonexpansive*.

Proximal Point Algorithm (PPA) [Roc76]

Starting from $x_0 \in \mathcal{H}$, iterate

$$x_{k+1} \in \operatorname{prox}_{\gamma_k f}(x_k), \qquad k = 0, 1, 2, \ldots$$

with step sizes $\gamma_k > 0$.

Interpretation as a **fixed-point iteration**.

Convex f:

$$x_{k+1} = \operatorname{prox}_{\gamma_k f}(x_k), \qquad k = 0, 1, 2, \ldots$$

$$x^\star \in \arg\min f \iff x^\star \in \operatorname{prox}_{\gamma f}(x^\star) \iff 0 \in \partial f(x^\star)$$

In general, there is a hierarchy [TSP18]

$$x^\star \in \arg\min f \implies \exists \gamma > 0\colon x^\star \in \operatorname{prox}_{\gamma f}(x^\star) \implies 0 \in \partial f(x^\star)$$

optimality $\implies$ γ-criticality $\implies$ stationarity

Interpretations: Operator theory

Consider f **convex** and its **subdifferential operator** $\hat{\partial} f$:

$$\hat{\partial} f(x) \coloneqq \{y \mid f(z) \geq f(x) + y^\top (z - x) \text{ for all } z \in \operatorname{dom} f\}.$$

The point-to-set mapping ∂f takes each point $x \in \operatorname{dom} f$ to the set $\partial f(x)$.
Examples: $f \coloneqq |\cdot|$ and $f \coloneqq \delta_{[-1,1]}$.

Any point $y \in \partial f(x)$ is called a *subgradient* of f at x. When f is differentiable, we have $\partial f(x) = \{\nabla f(x)\}$ for all x.

We use the limiting (Mordukhovich) subdifferential ∂f when f is nonconvex. For convex f, $\partial f = \hat{\partial} f$. Example: $f \coloneqq -|\cdot|$.

Characterization for **convex** f (subdifferential and Fermat):

$$x = \operatorname{prox}_{\gamma f}(y) = \operatorname*{arg\,min}_z \left\{ f(z) + \frac{1}{2\gamma} \|z - y\|^2 \right\}$$

$$\iff 0 \in \partial f(x) + \frac{x - y}{\gamma} \iff x = (\mathbb{I} + \gamma \partial f)^{-1}(y).$$

Monotone **operator theory:** $(\mathbb{I} + \gamma \partial f)^{-1}$ is called the **resolvent** of $\gamma \partial f$.

Monotone Operator Viewpoint

For **convex** f, the **subdifferential** ∂f is a *maximally monotone operator*:

$$\langle u - v, x - y \rangle \geq 0 \quad \forall\, (x, u), (y, v) \in \operatorname{gph}(\partial f).$$

The proximal step is the **resolvent** of ∂f:

$$J_{\lambda \partial f} := (\mathbb{I} + \lambda \partial f)^{-1} = \operatorname{prox}_{\lambda f}\,.$$

Hence the PPA solves the **inclusion** $x \in \operatorname{zer}(T)$, namely

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in T(x),$$

for a maximally monotone operator $T \equiv \partial f$, by iterating

$$x_{k+1} = J_{\lambda_k T}(x_k).$$

No convexity of f is required — only monotonicity of T, or less! [EPLP25]

Key Properties of the Resolvent

Let T be maximally monotone. The resolvent $J_{\lambda T} := (\mathbb{I} + \lambda T)^{-1}$ satisfies:

- **Single-valuedness** and **full domain**: $J_{\lambda T} : \mathcal{H} \to \mathcal{H}$ is well-defined everywhere.
- **Firm nonexpansiveness**:

$$\|J_{\lambda T}(x) - J_{\lambda T}(y)\|^2 \leq \langle J_{\lambda T}(x) - J_{\lambda T}(y), x - y \rangle.$$

- Firm nonexpansiveness $\implies$ **nonexpansiveness**:

$$\|J_{\lambda T}(x) - J_{\lambda T}(y)\| \leq \|x - y\|.$$

- **Cayley–Minty theorem**: $\mathrm{Fix}(J_{\lambda T}) = \mathrm{zer}(T)$.

Ryu & Boyd's "Primer on Monotone Operator Methods" [RB16]

Convergence: Exact Steps

Theorem [Roc76]

Let T be maximally monotone and $\text{zer}(T) \neq \emptyset$. If $\sum_{k=0}^{\infty} \lambda_k = +\infty$, then the sequence (x_k) generated by

$$x_{k+1} = J_{\lambda_k T}(x_k)$$

converges *weakly* to a point $x^\star \in \text{zer}(T)$.

Remarks:

- Weak convergence is tight in infinite dimensions (strong convergence can fail).
- **Strong convergence** holds if T^{-1} is *single-valued* (e.g., f is strictly convex).
- The condition $\sum \lambda_k = \infty$ prevents step sizes from vanishing too fast.

Convergence Rates and Inexactness

Sublinear convergence in general, even for convex f.

Under stronger assumptions

- **Strong monotonicity**/convexity: *linear* convergence $\|x^k - x^\star\| \le \rho^k \|x^0 - x^\star\|$.
- **Polyak–Łojasiewicz / error bounds**: local superlinear, or even *quadratic* convergence, with geometrically growing λ_k.

Inexact PPA. [Roc76] allows summable errors $\|e^k\|$:

$$x^{k+1} \approx J_{\lambda_k T}(x^k), \qquad \sum_{k=0}^{\infty} \|e^k\| < \infty$$

where $e^k := x^{k+1} - J_{\lambda_k T}(x^k)$.

Weak **convergence is preserved**. Many other criteria [Luq84, SS00, SV12].

Error Bounds: A short detour

Lourenço's "Error Bounds and Facial Residual Functions for Conic Linear Programs", OWOS 2022

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad h(x) = 0$$

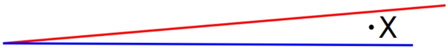
Imagine our favourite solver returns $x^\star$, declaring that the stationarity conditions are satisfied to within $\varepsilon = 10^{-6}$ and that $\|h(x^\star)\| \leq 10^{-7}$.

Questions

Is $x^\star$ close to the set of **optimal** solutions? Is $x^\star$ close to the set of **feasible** solutions?

Not necessarily!

forward error $\neq \mathcal{O}$(backward error)



- Solvers can only compute the residuals, $\|h(x^\star)\|$ for instance, (**backward error**)
- but true distance to the feasible region is hardly computable! (**forward error**)

Error bounds provide relations between forward and backward errors.

Related to metric (sub)regularity and Kurdyka–Łojasiewicz property [ABRS10, ABS13]

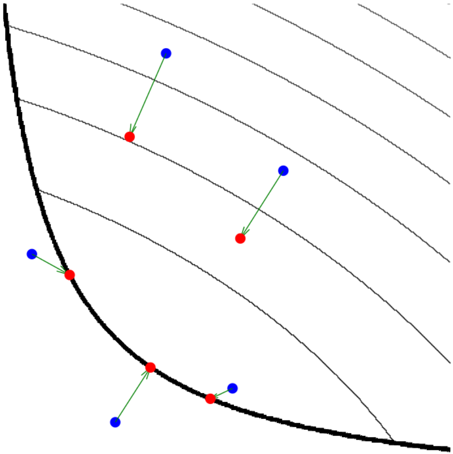
Interpretation: Proximal Minimization

Evaluating $\text{prox}_{\gamma f}$ at points x moves them to the corresponding z points.

x in the domain of f stay in the domain and move towards the minimum of the function.

The other two move to the boundary of the domain and towards the minimum of the function.
The parameter γ controls the trade-off between

- points towards the minimum of f
- and points closer to x.



Parikh & Boyd's [PB14]

Interpretation: Trust Region

Given some $v \in \mathbb{R}^n$, a **trust region** (TR) subproblem has the form

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \|x - v\| \leq \Delta$$

where $\Delta > 0$ is the TR radius. Typically, f is an approximation/surrogate/model for some true objective. Larger radii Δ reflect confidence in our model (around v).

The **proximal subproblem**

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + \frac{1}{2\gamma} \|x - v\|^2$$

is not too different:

- same two functions of x, f and $\|x - v\|$,
- but the TR constraint becomes a proximity-promoting penalty.

Roughly speaking, same solutions for appropriate choices of Δ and γ [PB14]

Interpretation: Discretized dynamical system

The simplest algorithm to solve a **smooth** optimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x)$$

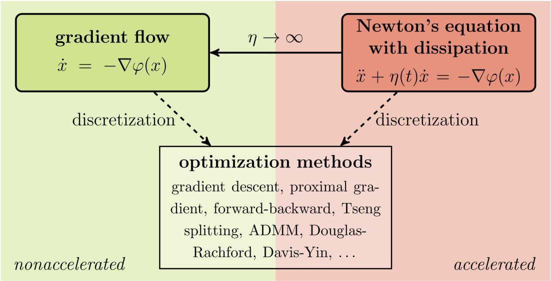
dates back to Cauchy (1847). It is the well-known **gradient descent** method:

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

where $\gamma_k > 0$ is the step size. It is the **explicit** Euler discretization of the gradient flow:

$$\dot{x} = -\nabla f(x).$$

“Gradient flows and proximal splitting methods” [FRV21]:
Several well-known optimization methods arise as dicretizations of the gradient flow. [...] Therefore, these *optimization methods consist of actual simulations of a classical dissipative physical system*.



Interpretation: Discretized dynamical system

The **nonsmooth** optimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x)$$

leads to the *differential inclusion*

$$\dot{x} \in -\partial f(x).$$

The **implicit** Euler discretization yields

$$x_{k+1} - x_k \in -\gamma_k \partial f(x_{k+1})$$
$$\iff \quad 0 \in \partial f(x_{k+1}) + \frac{x_{k+1} - x_k}{\gamma_k} \quad \Longleftarrow \quad x_{k+1} \in \operatorname{prox}_{\gamma_k f}(x_k)$$

Proximal methods are also called **backward schemes** [Ban14]

Nonsmooth (nonconvex) examples

Nonsmooth convex examples

Let us minimize $\|x\|_1$ and $\|x\|$.

First, compute prox of ℓ_1 and ℓ_2.

$$\text{prox}_{\gamma|\cdot|}(x) = \begin{cases} 0 & \text{if } |x| \leq \gamma, \\ x - \text{sign}(x)\gamma & \text{if } |x| \geq \gamma. \end{cases}$$

$$\text{prox}_{\gamma\|\cdot\|}(x) = \begin{cases} 0 & \text{if } \|x\| \leq \gamma, \\ \left(1 - \frac{\gamma}{\|x\|}\right) x & \text{if } \|x\| \geq \gamma. \end{cases}$$

[link to tutorial: PPA]

Nonsmooth nonconvex example

Let us minimize $\|x\|_0$.

First, compute prox of ℓ_0.

$$\mathrm{prox}_{\gamma|\cdot|_0}(x) = \begin{cases} 0 & \text{if } |x| < \sqrt{2\gamma} \\ \{0, x\} & \text{if } |x| = \sqrt{2\gamma} \\ x & \text{if } |x| > \sqrt{2\gamma} \end{cases}$$

[link to tutorial: PPA]

Solving QPs with PPA

Quadratic programming

What are **convex** QPs?

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}x^\top Qx + x^\top q \quad \text{subject to} \quad \ell \leq Ax \leq u$$

where $Q \succeq 0$, $A \in \mathbb{R}^{m\times n}$, $\ell, u \in \overline{\mathbb{R}}^m$, $\ell \leq u$

Applications: regression, portfolio optimization, linear MPC, optimal transport, ...

Recent works/solvers **involving proximal methods**:

- QPNNLS PROX [Bem18]
- OSQP [SBG⁺20]
- FBstab [LMK20]
- QPALM [HTP19, HTP22], QPDO [DM21], ProxQP [BSEK⁺25]
- IP-PMM [PG21], PS-IPM [CG23], PIQP [SJKJ23], Odyn [RPM⁺26]
- complexity/certification [ABA20, KITB24]

Equality-constrained QP

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \underbrace{\frac{1}{2} x^\top Q x + x^\top q}_{=: f(x)} \quad \text{subject to} \quad Ax = b$$

Lagrangian function

$$\mathscr{L}(x, y) := f(x) + y^\top (Ax - b)$$

KKT conditions

$$0 = \begin{pmatrix} \nabla_x \mathscr{L}(x, y) \\ -\nabla_y \mathscr{L}(x, y) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + A^\top y \\ b - Ax \end{pmatrix} =: \mathcal{T}(x, y)$$

Just a linear system, so easy and yet nontrivial!

$$\begin{bmatrix} Q & A^\top \\ A & \cdot \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

Why?

- non-strict convexity: Q not positive definite.
- redundant constraints: A with linearly dependent rows.

What would PPA do? Why would it help? Inherent regularization

Let's **solve the inclusion** $0 \in \mathcal{T}(x,y)$ **with PPA**. Denote $z \equiv (x,y)$.

$$\text{resolvent:} \qquad J_{\gamma\mathcal{T}} := (\mathbb{I} + \gamma\mathcal{T})^{-1} \quad \approx \text{prox}_{\gamma QP}$$

$$\text{PPA:} \qquad z_{k+1} = J_{\gamma_k \mathcal{T}}(z_k)$$

$$(\mathbb{I} + \gamma_k \mathcal{T})(z) = z_k \quad \Longleftrightarrow \quad \mathcal{T}(z) + \frac{z - z_k}{\gamma_k} = 0$$

$$\Longleftrightarrow \quad \begin{pmatrix} \nabla f(x) + A^\top y \\ b - Ax \end{pmatrix} + \frac{1}{\gamma_k} \begin{pmatrix} x - x_k \\ y - y_k \end{pmatrix} = 0$$

$$\Longleftrightarrow \quad \begin{bmatrix} Q + \frac{1}{\gamma_k}\mathbb{I} & A^\top \\ A & -\frac{1}{\gamma_k}\mathbb{I} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix} + \begin{bmatrix} \frac{1}{\gamma_k}\mathbb{I} & \cdot \\ \cdot & -\frac{1}{\gamma_k}\mathbb{I} \end{bmatrix} \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

for all $Q \succeq 0$, $A \in \mathbb{R}^{m \times n}$ and $\gamma_k > 0$.

The LHS matrix is **SQD**, hence the linear system always admits a unique solution.

[Vanderbei 1994]

PPA on EQP = **iterative refinement** for solving the (unregularized) linear system.

EQP examples: when everything is all right

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}(x_1 + x_2)^2 \quad \text{subject to} \quad x_1 - x_2 = 0$$

Unique solution: $x^\star = 0$; $y^\star = 0$.

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \end{pmatrix}$$

Linear system

$$\begin{bmatrix} Q & A^\top \\ A & \cdot \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

$\implies$ unique solution $x = (0, 0)$, $y = 0$.

[link to tutorial: QP]

EQP examples: redundant constraints

Same objective, double the constraints:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}(x_1 + x_2)^2 \quad \text{subject to} \quad x_1 - x_2 = 0, \quad x_1 - x_2 = 0$$

Solution: $x^\star = 0$; $y^\star \in \mathbb{R}^2$ s.t. $y_1^\star + y_2^\star = 0$.

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Linear system

$$\begin{bmatrix} Q & A^\top \\ A & \cdot \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

$\implies$ unique solution $x = (0, 0)$, non-unique y s.t. $y_1 + y_2 = 0$.

EQP examples: non-strict convexity

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}(x_1 - x_2)^2 \quad \text{subject to} \quad x_1 - x_2 = 0$$

Solution: any $x^\star \in \mathbb{R}^2$ s.t. $x_1^\star = x_2^\star$; $y^\star = 0$.

NOTE: Previous examples were also non-strictly convex, but the constraint saved the day there.

$$Q = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad A = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad b = \begin{pmatrix} 0 \end{pmatrix}$$

Linear system

$$\begin{bmatrix} Q & A^\top \\ A & \cdot \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

$\implies$ non-unique x s.t. $x_1 = x_2$, unique solution $y = 0$.

General convex QPs

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \underbrace{\frac{1}{2}x^\top Qx + x^\top q}_{=:f(x)} \quad \text{subject to} \quad Ax \in [\ell, u] =: C$$

$$\iff \quad \underset{x\in\mathbb{R}^n, w\in\mathbb{R}^m}{\text{minimize}} \quad f(x) + \delta_C(w) \quad \text{subject to} \quad Ax - w = 0$$

Seek to solve the KKT inclusion via PPA. What is the proximal subproblem?

A regularized QP:

$$\begin{aligned} \underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad & f(x) + \frac{1}{2\gamma_k}\|x - x_k\|^2 + \frac{\gamma_k}{2}\left\|s - \frac{y_k}{\gamma_k}\right\|^2 \\ \text{subject to} \quad & \ell \leq Ax + s \leq u \end{aligned}$$

A triumph of regularization: always feasible, linearly independent constraints, strictly convex.

Solvers exploit this! [HTP19, LMK20, DM21]

Connection to ALM

Equality-constrained NLP

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) = 0 \qquad \text{(ENLP)}$$

PPA subproblem for KKT inclusion: find (x, y) such that

$$\begin{pmatrix} \nabla f(x) + J_c(x)^\top y \\ -c(x) \end{pmatrix} + \begin{bmatrix} \frac{1}{\gamma_k} & \cdot \\ \cdot & \frac{1}{\gamma_k} \end{bmatrix} \begin{pmatrix} x - x_k \\ y - y_k \end{pmatrix} = 0$$

$$\iff \quad \begin{cases} 0 = \nabla f(x) + J_c(x)^\top y + \frac{1}{\gamma_k}(x - x_k) \\ 0 = c(x) + \frac{1}{\gamma_k}(y_k - y) \end{cases}$$

Formally solving for $y = y_k + \gamma_k c(x)$, it remains to solve

$$0 = \nabla f(x) + J_c(x)^\top [y_k + \gamma_k c(x)] + \frac{1}{\gamma_k}(x - x_k)$$

The RHS can be seen as the gradient of

$$\mathscr{L}_{\gamma_k}(x; x_k, y_k) := f(x) + y_k^\top c(x) + \frac{\gamma_k}{2}\|c(x)\|^2 + \frac{1}{2\gamma_k}\|x - x_k\|^2$$

which is the **proximal augmented Lagrangian function** of (ENLP)

PPA: Splitting methods and more

Proximal point methods: **high-level framework, abstract and flexible**
Very active area of research, especially in the convex setting (deep theory, strong guarantees)

Interesting even for simple problems: convex QP.

More general problems: **conic and semidefinite programming**. Solvers `SCS` [OCPB16], `ProxSDP` [SGV20], `COSMO` [GCG21], `Clarabel` [GC24].

Generic acceleration scheme `SuperMann` [TP19] (Anderson, BFGS, . . .).

Characterization for **convex composite problems** [DMHM25, ACT25]:

$$\underset{x}{\text{minimize}}\, f(x) + g(c(x)).$$

Gradient-antigradient flow for NLP [BGPG20, PB21]

Still active area of research

[LT25]: "We study the proximal point algorithm when the operator of interest is metrically subregular and satisfies a **submonotonicity** property. The latter property can be viewed as a quantified weakening of the standard definition of a monotone operator. Our main result gives a condition under which **locally**, the proximal point algorithm generates sequences that are linearly convergent to a zero of the underlying operator."

[ACT25]: "This work investigates the convergence behavior of augmented Lagrangian methods when applied to convex optimization **problems that may be infeasible**. [...] This study leverages the classical relationship between ALMs and the proximal-point algorithm applied to the dual problem. A key technical contribution is a set of concise results on the behavior of the proximal-point algorithm when applied to **functions that may not have minimizers**."

[DMHM25]: "This paper is concerned with augmented Lagrangian methods for the treatment of fully convex composite optimization problems. We extend the classical relationship between augmented Lagrangian methods and the proximal point algorithm to the **inexact and safeguarded scheme** in order to state global primal-dual convergence results. Our analysis distinguishes the regular case, where a stationary minimizer exists, and the **irregular case, where all minimizers are nonstationary**."

References I

Daniel Arnström, Alberto Bemporad, and Daniel Axehill.
Complexity certification of proximal-point methods for numerically stable quadratic programming.
IEEE Control Systems Letters, 5(4):1381–1386, 2020.

Hedy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran.
Alternating minimization and projection methods for structured nonconvex problems.
Mathematics of Operations Reserch, 35(2):438–457, 2010.

Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter.
Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods.
Mathematical Programming, 137(1):91–129, 2 2013.

Roland Andrews, Justin Carpentier, and Adrien Taylor.
Augmented Lagrangian methods for infeasible convex optimization problems and diverging proximal-point algorithms, 2025.

Sebastian Banert.
Backward–backward splitting in Hadamard spaces.
Journal of Mathematical Analysis and Applications, 414(2):656–665, 2014.

References II

Alberto Bemporad.
A numerically stable solver for positive semidefinite quadratic programs based on nonnegative least squares.
IEEE Transactions on Automatic Control, 63(2):525–531, 2 2018.

Hans Georg Bock, Jürgen Gutekunst, Andreas Potschka, and María Elena Suaréz Garcés.
A flow perspective on nonlinear least-squares problems.
Vietnam Journal of Mathematics, 48(4):987–1003, 12 2020.

Antoine Bambade, Fabian Schramm, Sarah El-Kazdadi, Stéphane Caron, Adrien Taylor, and Justin Carpentier.
Proxqp: an efficient and versatile quadratic programming solver for real-time robotics applications and beyond.
IEEE Transactions on Robotics, pages 1–19, 2025.

Stefano Cipolla and Jacek Gondzio.
Proximal stabilized interior point methods and low-frequency-update preconditioning techniques.
Journal of Optimization Theory and Applications, 197(3):1061–1103, 2023.

Alberto De Marchi.
QPDO: the Quadratic Primal-Dual Optimizer, 2021.

References III

Alberto De Marchi, Tim Hoheisel, and Patrick Mehlitz.
Augmented Lagrangian methods for fully convex composite optimization, 2025.

Brecht Evens, Pieter Pas, Puya Latafat, and Panagiotis Patrinos.
Convergence of the preconditioned proximal point method and douglas–rachford splitting in the absence of monotonicity.
Mathematical Programming, 214(1):247–301, 2025.

Guilherme França, Daniel P. Robinson, and René Vidal.
Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization.
Phys. Rev. E, 103:053304, May 2021.

Paul J. Goulart and Yuwen Chen.
Clarabel: An interior-point solver for conic programs with quadratic objectives, 2024.

Michael Garstka, Mark Cannon, and Paul Goulart.
COSMO: A conic operator splitting method for convex conic problems.
Journal of Optimization Theory and Applications, 190(3):779–810, 2021.

Ben Hermans, Andreas Themelis, and Panagiotis Patrinos.
QPALM: A Newton-type proximal augmented Lagrangian method for quadratic programs.
In *IEEE 58th Conference on Decision and Control (CDC)*, pages 4325–4330, Nice, France, 12 2019.

References IV

Ben Hermans, Andreas Themelis, and Panagiotis Patrinos.
QPALM: A proximal augmented Lagrangian method for nonconvex quadratic programs.
Mathematical Programming Computation, 14(3):497–541, 2022.

P. Krupa, O. Inverso, M. Tribastone, and Alberto Bemporad.
Certification of the proximal gradient method under fixed-point arithmetic for box-constrained qp problems.
Automatica, 159, 2024.

Dominic Liao-McPherson and Ilya Kolmanovsky.
FBstab: A proximally stabilized semismooth algorithm for convex quadratic programming.
Automatica, 113:108801, 2020.

D. Russell Luke and Matthew K. Tam.
Generalized monotonicity and the proximal point algorithm.
Mathematics of Operations Research, 2025.
online, ahead of print.

Fernando Javier Luque.
Asymptotic convergence analysis of the proximal point algorithm.
SIAM Journal on Control and Optimization, 22(2):277–293, 1984.

References V

Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd.
Conic optimization via operator splitting and homogeneous self-dual embedding.
Journal of Optimization Theory and Applications, 169(3):1042–1068, 6 2016.

Neal Parikh and Stephen Boyd.
Proximal algorithms.
Foundations and Trends in Optimization, 1(3):127–239, 2014.

Andreas Potschka and Hans Georg Bock.
A sequential homotopy method for mathematical programming problems.
Mathematical Programming, 187(1):459–486, 2021.

Spyridon Pougkakiotis and Jacek Gondzio.
An interior point-proximal method of multipliers for convex quadratic programming.
Computational Optimization and Applications, 78(2):307–351, 2021.

Ernest K. Ryu and Stephen Boyd.
A primer on monotone operator methods.
Applied and Computational Mathematics, 15(1):3–43, 2016.

Ralph Tyrrell Rockafellar.
Monotone operators and the proximal point algorithm.
SIAM Journal on Control and Optimization, 14(5):877–898, 1976.

References VI

Jose Rojas, Aristotelis Papatheodorou, Sergi Martinez, Ioannis Havoutis, and Carlos Mastalli.
ODYN: An all-shifted non-interior-point method for quadratic programming in robotics and AI, 2026.

B. Stellato, G. Banjac, P. Goulart, A. Bemporad, and S. Boyd.
OSQP: an operator splitting solver for quadratic programs.
Mathematical Programming Computation, 12(4):637–672, 2020.

Mario Souto, Joaquim D. Garcia, and Álvaro Veiga.
Exploiting low-rank structure in semidefinite programming by approximate operator splitting.
Optimization, pages 1–28, 2020.

Roland Schwan, Yuning Jiang, Daniel Kuhn, and Colin N. Jones.
PIQP: A proximal interior-point quadratic programming solver.
In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 1088–1093, 2023.

M. V. Solodov and B. F. Svaiter.
An inexact hybrid generalized proximal point algorithm and some new results on the theory of bregman functions.
Mathematics of Operations Research, 25(2):214–230, 2000.

Saverio Salzo and Silvia Villa.
Inexact and accelerated proximal point algorithms.
Journal of Convex analysis, 19(4):1167–1192, 2012.

References VII

Andreas Themelis and Panagiotis Patrinos.
SuperMann: A superlinearly convergent algorithm for finding fixed points of nonexpansive operators.
IEEE Transactions on Automatic Control, 64(12):4875–4890, 12 2019.

Andreas Themelis, Lorenzo Stella, and Panagiotis Patrinos.
Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms.
SIAM Journal on Optimization, 28(3):2274–2303, 2018.