

Binding-Set Stability for Quadratic Option Market-Making Approximations*

Aryan Ayyar

Abstract

Quadratic approximations are widely used to compute smooth quote maps in high-dimensional option market making. Their constrained implementation, however, depends on a discrete object that the smooth quote map does not control: the set of option sides whose optimal quotes bind at upper operational caps. This paper studies that binding set. Conditional on the HJB pressure vector generated by the option book, a scalar inverse-hazard cap condition induces an exact state-dependent diagnostic. The main result compares this exact diagnostic with the one obtained from an approximate, for example quadratic, pressure field. If the pressure approximation error is $\varepsilon(z) = \|P^Q(z) - P(z)\|_\infty$, then approximate and exact binding sets can disagree only on sides whose exact pressures lie within $\varepsilon(z)$ of their cap thresholds. Thus the approximate diagnostic is exact whenever the approximation error is smaller than the margin to the nearest threshold. Operationally, this margin is a stress indicator: it tells a quoting engine when projected quadratic quotes are reliable and when the cap decision is intrinsically fragile. An observable-margin corollary replaces the unavailable exact margin by the computable approximate margin under a certified pressure-error bound. The paper also records the monotone up-set structure of the binding map, connects value-function approximation error to pressure error, and shows how heterogeneous or mixed-sign Greek loadings generate non-nested partial-binding patterns.

1. Introduction

Modern option market makers quote many related instruments at once. A single fill changes not only the inventory of one contract, but also the maker's exposure to common risk factors:

*Preliminary draft: May 31, 2026.

Delta, Vega, Gamma, maturity buckets, smile risk, and other shared Greeks. Because these risks are shared across the book, the cost of quoting one option side depends on the inventory and risk pressure created by many other option positions. This coupling is exactly why high-dimensional market-making models are hard. The full Hamilton–Jacobi–Bellman equation contains value-function differences across a large inventory state space, and direct numerical solution quickly becomes impractical. Quadratic approximations are therefore attractive: they replace the high-dimensional value function by a tractable local approximation and produce smooth quote adjustments that can be computed efficiently.

The scalar cap condition used below is not the main novelty. It is the upper-endpoint version of the familiar inverse-hazard first-order condition for a one-sided market-making problem. Once the HJB has produced a pressure vector and fills are controlled side by side, each side solves such a scalar problem. The contribution is the book-level binding-set diagnostic induced by those scalar thresholds, and the stability result that says when an approximate pressure field recovers that diagnostic correctly.

The central object is the binding set: the set of option sides whose effective pressures exceed their side-specific upper-cap thresholds. In a scalar model this object is trivial. In an option book it is a state-dependent subset of legs and sides. Different options load differently, and sometimes with different signs, on shared risk factors. Their effective pressures therefore need not approach their thresholds in nested order. A stressed book can display partial binding: some sides are capped while others remain governed by the interior quote map.

The paper makes three contributions. First, it states the sidewise upper-cap condition and its monotone comparative-static structure. The binding map is an up-set in pressure space: raising any pressure cannot remove a side that was already binding. Second, it gives option-book examples in which heterogeneous and mixed-sign Greek loadings generate strict partial-binding regions. This frames partial binding as a generic book phenomenon rather than as a scalar threshold curiosity. Third, and most importantly, it proves a binding-set stability theorem for approximate pressures. If P is the exact pressure vector, P^Q is the quadratic pressure vector, and $\varepsilon = \|P^Q - P\|_\infty$, then the approximate and exact binding sets can differ only on sides whose exact threshold margin is at most ε . This is deliberately a classifier-stability theorem, not a new smooth HJB approximation theorem. Its point is

that the implementation risk is discrete: a small pressure error matters only when a side is already near an upper-cap switching surface.

The analysis is closest to high-dimensional option market-making approximations such as Baldacci et al. (2020b), option-specific stochastic-control models such as Baldacci et al. (2020a), and recent adverse-selection and price-reading models such as Barzykin et al. (2025). Existing papers approximate smooth quote maps or perturb interior quoting equations. This paper studies the constrained implementation layer: the state-dependent set on which the smooth quote ceases to be the constrained optimizer, and the conditions under which an approximate pressure field identifies that set correctly.

2. Related Literature

This paper sits at the intersection of inventory-based market making, high-dimensional approximation methods, option-specific quoting, and constrained implementation. The classical microstructure lineage separates inventory pressure from informational pressure. Ho and Stoll (1981) study dealer pricing under inventory and return uncertainty, while Glosten and Milgrom (1985) and Kyle (1985) provide foundational models of informed trading and price discovery. Modern stochastic-control market-making models, including Avellaneda and Stoikov (2008), Guéant et al. (2013), Guilbaud and Pham (2013), Cartea et al. (2014), and the monograph of Cartea et al. (2015), make inventory-sensitive quoting mathematically tractable and operationally interpretable.

High-dimensional market making raises a separate approximation problem. Guéant (2017) gives a broad optimal market-making framework with multi-asset approximations, and Bergault et al. (2021) develop closed-form approximations for multi-asset market making. In the option setting, Baldacci et al. (2020b) develop approximate methods for options market making in high dimension, while Baldacci et al. (2020a) provide an option-specific framework in which option inventories and Greeks are central state variables. These papers show why smooth approximations are useful: they turn otherwise intractable HJB systems into computationally usable quote-adjustment rules. The stability result below asks a different question: when does an approximate pressure field recover the same constrained binding set as the exact pressure field?

The closest recent adverse-selection and price-reading paper is Barzykin et al. (2025). That paper studies adverse selection and price reading in an optimal-quoting framework and derives first-order effects on value functions and optimal quotes, including under quadratic Hamiltonian approximations. Its contribution is interior and perturbative: informational pressure changes smooth quote corrections. The present paper studies the boundary of that interior regime after upper caps or projection rules are imposed. The wedge is useful only if the binding set carries structure; the results below provide that structure through up-set monotonicity, partial-binding geometry, and approximation stability.

Option market making adds a Greek-inventory layer that is absent from scalar market-making models. El Aoud and Abergel (2015) study option market making through stochastic control under volatility-model misspecification, and Lucic and Tse (2024) emphasize option quoting with volatility-arbitrage considerations. Here the side-specific binding set $\mathcal{M}(t, x, q, A)$ records which legs or sides remain interior and which must be assigned to an upper cap, projection, or a separate no-quote action. Because option legs can have heterogeneous and mixed-sign Greek loadings, the binding set is not generally obtained by sorting loading magnitudes. Its geometry depends on pressure signs, threshold margins, caps, and toxicity shifts.

3. Binding-Set Stability for Quadratic Approximations

This section states the main approximation result before the full market-making model is introduced. Let \mathcal{K} be a finite set of side-level controls, let $P(z) = (P_k(z))_{k \in \mathcal{K}}$ be the exact pressure vector at state $z = (t, x, q, A)$, and let $B_k(A)$ be the side-specific cap threshold. The model in Section 4 derives these objects from the option-book HJB and the scalar cap condition.

A computed quote engine rarely has the exact pressure vector. In high-dimensional option books the pressure field is typically supplied by a quadratic value-function approximation, a reduced Greek-factor model, or another local surrogate. The relevant question is therefore not only whether the smooth approximate quote is close to the exact smooth quote. The constrained implementation depends on a discrete object,

$$\mathcal{M}(z) \subseteq \mathcal{K}, \quad z = (t, x, q, A),$$

and small pressure errors can matter only if they move a side across its threshold.

The result below is intentionally elementary in form. It is a stability theorem for a threshold classifier under pressure perturbations. The contribution is not that threshold classifiers are hard. It is that the threshold classifier is the missing constrained-implementation layer in a quadratic option market-making engine: the continuous approximation produces pressures or smooth quotes, while the production system must decide which option sides have crossed upper operational caps.

Let $P^Q(z)$ denote the pressure vector produced by the quadratic approximation. It induces the approximate binding diagnostic

$$\mathcal{M}^Q(z) = \{ k \in \mathcal{K} : P_k^Q(z) \geq B_k(A) \},$$

while the exact diagnostic is

$$\mathcal{M}(z) = \{ k \in \mathcal{K} : P_k(z) \geq B_k(A) \}.$$

Define the exact and approximate signed threshold gaps

$$\Delta_k(z) = P_k(z) - B_k(A), \quad \Delta_k^Q(z) = P_k^Q(z) - B_k(A),$$

the pointwise pressure error

$$\varepsilon(z) = \|P^Q(z) - P(z)\|_\infty,$$

and the exact margin to the nearest threshold

$$\gamma(z) = \min_{k \in \mathcal{K}} |\Delta_k(z)|.$$

Also define the observable approximate margin

$$\gamma^Q(z) = \min_{k \in \mathcal{K}} |\Delta_k^Q(z)|.$$

The next result is the paper's main approximation statement. It says that the only fragile

states are the states already close to an exact threshold surface.

Theorem 1 (Binding-set stability under pressure error). *For every state z ,*

$$\mathcal{M}^Q(z) \Delta \mathcal{M}(z) \subseteq \{k \in \mathcal{K} : |\Delta_k(z)| \leq \varepsilon(z)\}.$$

Equivalently, if a side satisfies

$$|P_k(z) - B_k(A)| > \varepsilon(z),$$

then the quadratic and exact diagnostics assign that side to the same regime:

$$k \in \mathcal{M}^Q(z) \iff k \in \mathcal{M}(z).$$

Proof. Fix z and suppose $k \in \mathcal{M}^Q(z) \Delta \mathcal{M}(z)$. Then the exact and approximate signed gaps put side k on opposite sides of zero: one of $\Delta_k(z)$ and $\Delta_k^Q(z)$ is nonnegative and the other is negative, allowing equality at the threshold. Hence zero lies between them, so

$$|\Delta_k(z)| \leq |\Delta_k^Q(z) - \Delta_k(z)| = |P_k^Q(z) - P_k(z)| \leq \varepsilon(z).$$

This proves the inclusion. The equivalent statement is its contrapositive, side by side. \square

Corollary 1 (Margin recovery). *If*

$$\varepsilon(z) < \gamma(z),$$

then

$$\mathcal{M}^Q(z) = \mathcal{M}(z).$$

Proof. If the binding sets differed, Theorem 1 would give a side k with

$$|\Delta_k(z)| \leq \varepsilon(z).$$

But $|\Delta_k(z)| \geq \gamma(z)$ for every side k , contradicting $\varepsilon(z) < \gamma(z)$. \square

Corollary 2 (Observable-margin recovery). *Suppose the implementation has a certified pressure-error bound*

$$\bar{\varepsilon}(z) \geq \|P^Q(z) - P(z)\|_\infty.$$

If

$$\bar{\varepsilon}(z) < \gamma^Q(z),$$

then

$$\mathcal{M}^Q(z) = \mathcal{M}(z).$$

Proof. For each side k ,

$$|\Delta_k(z) - \Delta_k^Q(z)| = |P_k(z) - P_k^Q(z)| \leq \bar{\varepsilon}(z).$$

If $k \in \mathcal{M}^Q(z)$, then $\Delta_k^Q(z) \geq 0$. Since $|\Delta_k^Q(z)| \geq \gamma^Q(z) > \bar{\varepsilon}(z)$, it follows that $\Delta_k^Q(z) > 0$ and

$$\Delta_k(z) \geq \Delta_k^Q(z) - \bar{\varepsilon}(z) > 0,$$

so $k \in \mathcal{M}(z)$. If $k \notin \mathcal{M}^Q(z)$, then $\Delta_k^Q(z) < 0$ and $|\Delta_k^Q(z)| > \bar{\varepsilon}(z)$, so

$$\Delta_k(z) \leq \Delta_k^Q(z) + \bar{\varepsilon}(z) < 0,$$

and $k \notin \mathcal{M}(z)$. Thus every side receives the same exact and approximate classification. \square

Remark 1 (Why the approximate margin matters). Corollary 1 is the clean mathematical statement, but its margin $\gamma(z)$ depends on the unavailable exact pressure. Corollary 2 is the operational version: a quote engine can compute $\gamma^Q(z)$, and if it also carries a certified pressure-error envelope $\bar{\varepsilon}(z)$, then the approximate binding set is reliable whenever the observable margin dominates that envelope.

Corollary 3 (Threshold-tube localization). *Fix a state region D . For each side k , let*

$$\Sigma_k = \{z \in D : \Delta_k(z) = 0\}$$

be its exact threshold surface. Suppose that, on a neighborhood of Σ_k ,

$$\|\nabla_z \Delta_k(z)\| \geq m_k > 0.$$

Then the disagreement set

$$\{z \in D : \mathcal{M}^Q(z) \neq \mathcal{M}(z)\}$$

is contained in a union of side-specific tubes around the exact threshold surfaces. Locally, the tube around Σ_k has thickness of order $\varepsilon(z)/m_k$.

Proof. By Theorem 1, disagreement can occur only where

$$|\Delta_k(z)| \leq \varepsilon(z)$$

for at least one side k . Near a regular threshold surface Σ_k , the gradient lower bound gives a local normal coordinate in which the signed gap changes at speed at least m_k . Therefore the sublevel set $\{|\Delta_k| \leq \varepsilon\}$ is locally contained in a tube around Σ_k whose thickness is bounded, up to the coordinate constant, by $\varepsilon(z)/m_k$. Taking the union over sides gives the claim. \square

Proposition 1 (From value-function error to pressure error). *Let u be the exact continuation value and u^Q a quadratic or other approximate continuation value. For side k , define*

$$p_k(z) = u(z) - u(z + \zeta_k), \quad p_k^Q(z) = u^Q(z) - u^Q(z + \zeta_k),$$

where $z + \zeta_k$ denotes the post-fill inventory state with the other state coordinates held fixed. Suppose that, on the one-fill neighborhood of z ,

$$\max_{k \in \mathcal{K}} \max\{|u^Q(z) - u(z)|, |u^Q(z + \zeta_k) - u(z + \zeta_k)|\} \leq \eta(z),$$

and that the markout approximation satisfies

$$\|c^Q(A) - c(A)\|_\infty \leq \kappa(z).$$

Then the effective pressure fields $P_k = p_k + c_k$ and $P_k^Q = p_k^Q + c_k^Q$ obey

$$\|P^Q(z) - P(z)\|_\infty \leq 2\eta(z) + \kappa(z).$$

In particular, if markout costs are not approximated, then $\|P^Q(z) - P(z)\|_\infty \leq 2\eta(z)$.

Proof. For each side k ,

$$|p_k^Q(z) - p_k(z)| \leq |u^Q(z) - u(z)| + |u^Q(z + \zeta_k) - u(z + \zeta_k)| \leq 2\eta(z).$$

Adding the side-specific markout error gives

$$|P_k^Q(z) - P_k(z)| \leq |p_k^Q(z) - p_k(z)| + |c_k^Q(A) - c_k(A)| \leq 2\eta(z) + \kappa(z).$$

Taking the maximum over k proves the claim. □

Combined with Corollary 2, this gives a direct quadratic-approximation diagnostic. If a numerical scheme certifies a local value-function error $\eta(z)$ on one-fill neighboring states and a markout error $\kappa(z)$, then

$$2\eta(z) + \kappa(z) < \gamma^Q(z)$$

is sufficient for exact recovery of the binding set.

The theorem gives a sharp reliability criterion for constrained implementation. A quadratic engine may still compute an interior quote map and then project it, for example

$$\delta_k^{Q,\text{proj}}(z) = \min\{\delta_k^{Q,\text{int}}(z), \delta_{k,\text{max}}\},$$

but the projection layer is credible as a binding-set diagnostic only through the margin comparison above. Away from the threshold tubes, \mathcal{M}^Q recovers the exact constrained regime even though the value function and pressure field are approximate. Inside the tubes, instability is not a defect of the theorem; it is the correct geometry of a discrete regime classifier near its switching surface.

4. Model

The market maker quotes n option instruments. Each instrument has two quote sides,

$$s \in \{a, b\},$$

where a is the side on which the client buys the option from the market maker, and b is the side on which the client sells the option to the market maker. Thus an ask fill reduces the market maker's inventory in that option, while a bid fill increases it. Let

$$\mathcal{J} = \{1, \dots, n\}, \quad \mathcal{K} = \mathcal{J} \times \{a, b\}$$

denote the set of option legs and the set of quotable leg-sides. For a fixed side s , write

$$\mathcal{K}_s = \mathcal{J} \times \{s\}.$$

Binding is always a side-level event. If $\mathcal{M} \subseteq \mathcal{K}$ is a binding set, its projection onto legs on side s is

$$\mathcal{M}_s = \{j \in \mathcal{J} : (j, s) \in \mathcal{M}\}.$$

This distinction matters in an option book: the ask side of one option can bind while the bid side of the same option, or either side of another option, remains interior.

For each $k = (j, s) \in \mathcal{K}$, the control is a quote offset

$$\delta_k \in [0, \delta_{k, \max}].$$

The offset is measured relative to the model mid or indifference reference price, so larger δ_k is less aggressive. The upper endpoint $\delta_{k, \max}$ is an operational cap. Throughout the main analysis, the endpoint has positive intensity, so reaching it is cap binding, or boundary quoting, rather than literal withdrawal. Zero-intensity endpoints and explicit no-quote actions are separate implementation choices and are discussed in Section 10.

The diagnostic studied in this paper is the upper-cap diagnostic. The lower endpoint $\delta_k = 0$

is an aggressiveness or minimum-spread constraint, not the cap event analyzed below. A production quoting system may also monitor lower-bound binding, but that is a separate endpoint classifier with a different economic interpretation. All binding sets \mathcal{M} in the main text refer to upper-cap binding unless stated otherwise.

The state is

$$X_t = (S_t, \nu_t, q_t, A_t),$$

where S_t denotes underlying prices, ν_t volatility or stochastic-volatility factors, $q_t \in \mathbb{R}^n$ option inventories, and A_t an adverse-selection, toxicity, or informational-risk state. Write $x = (S, \nu)$ and keep A explicit, since toxicity also enters fill intensities and markout costs.

For each side $k = (j, s)$, let the fill intensity be

$$\lambda_k(\delta_k; A_t).$$

Assumption 1 (Fill intensities and endpoint cap). For each side $k \in \mathcal{K}$ and each fixed A , the intensity $\lambda_k(\cdot; A)$ is positive, C^1 , and strictly decreasing on $[0, \delta_{k,\max}]$, where $0 < \delta_{k,\max} < \infty$. Its hazard

$$h_k(\delta; A) = -\frac{\partial_\delta \lambda_k(\delta; A)}{\lambda_k(\delta; A)}$$

is positive, finite, and nondecreasing in δ . The cap-binding results in Sections 3–5 treat the mandatory-quote problem with no outside no-quote payoff.

Let $u(t, x, q, A)$ be the exact continuation value. For $k = (j, s)$, let e_j be the j -th unit vector and define the inventory jump

$$\zeta_{j,a} = -e_j, \quad \zeta_{j,b} = e_j.$$

The ordinary HJB pressure is the continuation-value loss caused by one fill:

$$p_k(t, x, q, A) = u(t, x, q, A) - u(t, x, q + \zeta_k, A).$$

This is the object in which the high-dimensional option-book coupling resides. It depends on inventories, Greek loadings, volatility factors, cross-option hedging possibilities, and

value-function differences. The final quote choice may be scalar side by side, but the pressure field that feeds it is not scalar.

Let

$$c_k(A) \geq 0$$

be the side-specific expected markout, pickoff, or toxic-flow cost paid upon a fill. The effective pressure is

$$P_k(t, x, q, A) = p_k(t, x, q, A) + c_k(A).$$

The decomposition below should be read conditionally on this pressure vector. All dynamic and cross-sectional structure has already been compressed into P .

Remark 2 (Pressure sign convention). The sign convention is chosen so that the side-local payoff can be written as $\delta_k - P_k$. Indeed,

$$\lambda_k(\delta_k; A) [\delta_k - c_k(A) + u(t, x, q + \zeta_k, A) - u(t, x, q, A)] = \lambda_k(\delta_k; A)(\delta_k - P_k(t, x, q, A)).$$

Execution-dependent state jumps, informational updates, and side-specific cashflow conventions can be absorbed into P_k without changing the side-local analysis.

Proposition 2 (Sidewise decomposition from independent fill controls). *Suppose that, conditional on the state, option-side fills arrive as point processes with intensities $\lambda_k(\delta_k; A)$, and that each offset δ_k affects only side k 's fill intensity and execution cashflow. Conditional on the effective pressures P_k , the execution Hamiltonian separates as*

$$\sup_{\delta \in \prod_{k \in \mathcal{K}} [0, \delta_{k, \max}]} \sum_{k \in \mathcal{K}} \lambda_k(\delta_k; A) \{\delta_k - P_k(t, x, q, A)\}.$$

The high-dimensional inventory and Greek coupling remains in the pressure vector P ; it is not a contemporaneous cross-control term in the final quote offsets.

Proof. For side k , a fill changes continuation value by

$$u(t, x, q + \zeta_k, A) - u(t, x, q, A) = -p_k(t, x, q, A).$$

Including the side-specific markout cost, the instantaneous contribution from that side is

$$\lambda_k(\delta_k; A)\{\delta_k - p_k(t, x, q, A) - c_k(A)\} = \lambda_k(\delta_k; A)\{\delta_k - P_k(t, x, q, A)\}.$$

The assumptions make these contributions additive across k , and the control set is a product of side-level intervals. Maximizing over that product gives the displayed Hamiltonian. \square

Assumption 2 (Sidewise terminal control separability). The separated Hamiltonian in Proposition 2 is the terminal quote-control Hamiltonian used below. It may be imposed directly or derived from the independent fill-control structure above.

Conditional on P_k , the side-local constrained quote problem is

$$\max_{0 \leq \delta_k \leq \delta_{k,\max}} \lambda_k(\delta_k; A) [\delta_k - P_k(t, x, q, A)]. \quad (4.1)$$

This scalar problem is the last step of the control calculation. The high-dimensional content enters through the state-dependent pressure vector to which the scalar threshold is applied side by side.

Define the side-specific endpoint hazard

$$h_{k,\max}(A) = h_k(\delta_{k,\max}; A)$$

and the cap-binding threshold

$$B_k(A) = \delta_{k,\max} - \frac{1}{h_{k,\max}(A)}.$$

Lemma 1 proves that this is the endpoint cap threshold for the scalar problem.

The binding set is

$$\mathcal{M}(t, x, q, A) = \{k \in \mathcal{K} : P_k(t, x, q, A) \geq B_k(A)\}. \quad (4.2)$$

When $\mathcal{M}(t, x, q, A)$ is empty, every side remains below its upper-cap threshold. When it is nonempty, the constrained optimizer places at least one side at its cap. The corresponding

leg set on side s is

$$\mathcal{M}_s(t, x, q, A) = \{j \in \mathcal{J} : (j, s) \in \mathcal{M}(t, x, q, A)\}.$$

The important book-level feature is that \mathcal{M} is an arbitrary subset of leg-sides, not merely the empty set or all of \mathcal{K} . Heterogeneous caps, intensities, markout costs, and Greek loadings can therefore create partial cap binding. If the implementation includes a no-quote action, withdrawal is determined by an additional outside-option comparison after this cap-binding diagnostic.

Remark 3 (Cap binding versus withdrawal). If $\lambda_k(\delta_{k,\max}; A) > 0$, the market maker still receives fills at the boundary. The event $k \in \mathcal{M}$ is therefore cap binding, not withdrawal. Withdrawal requires either a no-fill endpoint or an explicit no-quote action, as discussed below.

5. Scalar Cap Binding

This section proves the scalar primitive that will be applied side by side in the option book. It treats the mandatory-quote problem with positive endpoint intensity and no outside no-quote payoff. When a no-quote payoff is added, cap binding is replaced or followed by a separate withdrawal comparison. Consider a single quote side. The dealer chooses

$$\delta \in [0, \delta_{\max}].$$

For fixed adverse-selection state A and effective pressure P , define

$$H(P, A) = \max_{0 \leq \delta \leq \delta_{\max}} \lambda(\delta; A)(\delta - P).$$

Assume the scalar analogue of Assumption 1.

Lemma 1 (Scalar cap-binding threshold). *Define*

$$B(A) = \delta_{\max} - \frac{1}{h(\delta_{\max}; A)}.$$

Then the upper cap δ_{\max} is a global maximizer of the side-local Hamiltonian if and only if

$$P \geq B(A).$$

If the inequality is strict, the cap is the unique maximizer.

Proof. Let

$$F(\delta; P, A) = \lambda(\delta; A)(\delta - P).$$

Differentiating,

$$\partial_\delta F(\delta; P, A) = \lambda(\delta; A) \left[1 - h(\delta; A)(\delta - P) \right].$$

Define

$$r(\delta; A) = \delta - \frac{1}{h(\delta; A)}.$$

If h is differentiable, monotonicity of the hazard gives

$$\partial_\delta r(\delta; A) = 1 + \frac{\partial_\delta h(\delta; A)}{h(\delta; A)^2} \geq 1.$$

When the hazard is merely monotone, the same conclusion follows directly: if $\delta_2 > \delta_1$ and $h(\delta_2; A) \geq h(\delta_1; A)$, then

$$\delta_2 - \frac{1}{h(\delta_2; A)} \geq \delta_1 - \frac{1}{h(\delta_1; A)}.$$

Hence $r(\cdot; A)$ is increasing. Moreover,

$$\partial_\delta F(\delta; P, A) \geq 0 \iff P \geq r(\delta; A).$$

If $P \geq r(\delta_{\max}; A)$, then $P \geq r(\delta; A)$ for every $\delta \leq \delta_{\max}$, so F is increasing on the admissible interval and the cap is globally optimal. Conversely, if $P < r(\delta_{\max}; A)$, then the left derivative at the endpoint is negative:

$$\partial_\delta F(\delta_{\max}; P, A) < 0.$$

Therefore, for all sufficiently small $\varepsilon > 0$,

$$F(\delta_{\max} - \varepsilon; P, A) > F(\delta_{\max}; P, A),$$

so the cap cannot be optimal. Since $r(\delta_{\max}; A) = B(A)$, the result follows. \square

An exact comparison threshold is available without the monotone-hazard condition; see Appendix Appendix A. Monotone hazard is not needed for existence of a global cap-binding threshold; it is used here to collapse the global comparison into the endpoint expression $B(A)$.

6. Adverse-Selection Comparative Static

Let the effective pressure decompose as

$$P = p + c(A),$$

where p is the ordinary HJB inventory pressure and $c(A)$ is the expected markout or adverse-selection cost. By Lemma 1, cap binding occurs when

$$p + c(A) \geq \delta_{\max} - \frac{1}{h(\delta_{\max}; A)}.$$

Equivalently, the cap-binding threshold in ordinary pressure is

$$\bar{p}(A) = \delta_{\max} - \frac{1}{h(\delta_{\max}; A)} - c(A).$$

Proposition 3 (Ordinary-pressure threshold under toxicity). *Treat p as the ordinary-pressure coordinate and vary only the side-local functions $c(A)$ and $h(\delta_{\max}; A)$. Assume c and $h(\delta_{\max}; \cdot)$ are differentiable in A . Then*

$$\partial_A \bar{p}(A) = \frac{h_A(\delta_{\max}; A)}{h(\delta_{\max}; A)^2} - c_A(A).$$

Thus the cap-binding boundary moves inward in ordinary pressure as A rises if and only if

$$c_A(A) > \frac{h_A(\delta_{\max}; A)}{h(\delta_{\max}; A)^2}.$$

Proof. Differentiate

$$\bar{p}(A) = \delta_{\max} - \frac{1}{h(\delta_{\max}; A)} - c(A)$$

with respect to A . This gives

$$\partial_A \bar{p}(A) = \frac{h_A(\delta_{\max}; A)}{h(\delta_{\max}; A)^2} - c_A(A).$$

The boundary moves inward when the ordinary-pressure threshold decreases, i.e. when $\partial_A \bar{p}(A) < 0$. This is exactly the stated inequality. \square

Adverse selection has two side-local effects. It can increase the markout cost $c(A)$, which lowers the ordinary inventory pressure needed for cap binding. It can also reshape the fill curve through the endpoint hazard $h(\delta_{\max}; A)$. The boundary moves inward only when the markout-cost effect dominates the hazard-shape effect. This qualification matters: the model should not be read as claiming that toxicity always moves the boundary inward.

This proposition is a side-local comparative static. If A is a state variable in the full dynamic control problem, the HJB pressure $p_k(t, x, q, A)$ may also change with A . That dynamic channel is absorbed into the pressure field and is separate from the static boundary comparison above.

Proposition 4 (Toxicity-driven binding cascade). *Fix a state (t, x, q) and suppose, for each side k , the ordinary pressure $p_k(t, x, q)$ is held fixed while the toxicity state A shifts the side-specific boundary through $c_k(A)$ and $h_k(\delta_{k,\max}; A)$. Define*

$$\bar{p}_k(A) = \delta_{k,\max} - \frac{1}{h_k(\delta_{k,\max}; A)} - c_k(A).$$

If each $\bar{p}_k(A)$ is continuous and weakly decreasing in A , then the binding set

$$\mathcal{M}(A) = \{ k \in \mathcal{K} : p_k(t, x, q) \geq \bar{p}_k(A) \}$$

is weakly expanding in A : if $A_2 \geq A_1$, then

$$\mathcal{M}(A_1) \subseteq \mathcal{M}(A_2).$$

If the crossing levels

$$A_k^* = \inf\{A : p_k(t, x, q) \geq \bar{p}_k(A)\}$$

are distinct, then sides enter the binding set in the order of their crossing levels.

Proof. If $A_2 \geq A_1$ and $k \in \mathcal{M}(A_1)$, then

$$p_k(t, x, q) \geq \bar{p}_k(A_1) \geq \bar{p}_k(A_2),$$

so $k \in \mathcal{M}(A_2)$. This proves monotone expansion. Distinct crossing levels order the first toxicity values at which each side satisfies the binding inequality, giving the stated cascade. \square

This lifts the side-local comparative static into the book-level object. Toxicity need not force full-book withdrawal. It can produce an ordered cascade in which sides cross their cap thresholds at different toxicity levels.

The next section uses the scalar threshold side by side. The additional object is the HJB-generated pressure vector, whose components determine which option sides reach their thresholds first.

7. Partial Binding in an Option Book

We now return to the option book. Maintain the sidewise separability assumption from the model section: conditional on the exact pressure vector, the final instantaneous quote-control problem separates across option-sides. For each side $k \in \mathcal{K}$, define effective pressure

$$P_k(t, x, q, A) = p_k(t, x, q, A) + c_k(A)$$

and cap-binding threshold

$$B_k(A) = \delta_{k,\max} - \frac{1}{h_k(\delta_{k,\max}; A)}.$$

The binding set is

$$\mathcal{M}(t, x, q, A) = \{ k \in \mathcal{K} : P_k(t, x, q, A) \geq B_k(A) \}.$$

Proposition 5 (Monotone up-set structure). *Fix A and thresholds $B_k(A)$. For a pressure vector $P \in \mathbb{R}^{|\mathcal{K}|}$, define*

$$\mathcal{M}(P) = \{ k \in \mathcal{K} : P_k \geq B_k(A) \}.$$

If $P' \geq P$ componentwise, then

$$\mathcal{M}(P) \subseteq \mathcal{M}(P').$$

Equivalently, for each side k , the event $k \in \mathcal{M}(P)$ is an up-set in pressure space, and the binding correspondence is order-preserving.

Proof. If $k \in \mathcal{M}(P)$, then $P_k \geq B_k(A)$. Since $P'_k \geq P_k$, it follows that $P'_k \geq B_k(A)$, so $k \in \mathcal{M}(P')$. \square

Theorem 2 (Sidewise characterization of the cap-binding set). *Suppose Assumptions 1 and 2 hold. Conditional on the exact HJB pressure vector, there exists an optimal constrained selector satisfying*

$$\delta_k^*(t, x, q, A) = \delta_{k,\max}, \quad k \in \mathcal{M}(t, x, q, A),$$

while for $k \notin \mathcal{M}(t, x, q, A)$, the upper cap does not bind and any optimum lies in the interior when the interior first-order condition has a feasible solution. When the endpoint optimum is unique, every optimal selector binds at the cap for $k \in \mathcal{M}(t, x, q, A)$.

The characterization permits partial binding whenever, for some but not all sides,

$$P_k(t, x, q, A) \geq B_k(A).$$

Whether a given subset is reached depends on the pressure field generated by the HJB.

Proof. Conditional on the exact HJB pressure vector, the side-local optimization problem separates across sides:

$$\max_{0 \leq \delta_k \leq \delta_{k,\max}} \lambda_k(\delta_k; A) [\delta_k - P_k(t, x, q, A)].$$

Applying Lemma 1 side by side, side k binds at its upper cap if and only if

$$P_k(t, x, q, A) \geq B_k(A).$$

At equality the cap is a global maximizer; if the scalar maximizer is not unique, the statement is interpreted as an optimal-selector result. This is precisely the condition $k \in \mathcal{M}(t, x, q, A)$. If $k \notin \mathcal{M}(t, x, q, A)$, the endpoint derivative condition fails and the upper cap is not optimal. When the interior first-order condition admits a feasible solution, the optimum is interior.

Because the condition is side-specific, there is no requirement that all option-sides bind simultaneously. The theorem therefore gives a binding-set diagnostic: given a pressure vector, it identifies which side-local controls remain interior and which controls require projection or another boundary action. The two-option Vega example below gives an explicit region in which the binding set is strict and nonempty.

The high-dimensionality is not absent from this theorem. It is contained in the pressure vector $p_k(t, x, q, A)$, which is generated by the HJB and reflects inventory, Greeks, volatility factors, cross-option exposures, and value-function differences. \square

8. Greek-Loading Examples

Proposition 6 (Generic partial binding under heterogeneous loadings). *Fix one side s and suppose the effective pressure on leg j has the one-factor form*

$$P_{j,s}(G, A) = \alpha v_j G + c_{j,s}(A), \quad \alpha > 0.$$

Assume $v_j > 0$ for the legs under consideration and define the side-leg threshold

$$G_{j,s}^*(A) = \frac{B_{j,s}(A) - c_{j,s}(A)}{\alpha v_j}.$$

If the thresholds $G_{j,s}^*(A)$ are not all equal, then there is an open interval of G -values on which the side-specific leg binding set $\mathcal{M}_s(G, A)$ is a strict nonempty subset of \mathcal{J} . Thus simultaneous full-side binding is nongeneric: ties of threshold-crossing levels are codimension-one restrictions on loadings, costs, or caps.

Proof. Order the finite collection of thresholds $G_{j,s}^*(A)$. If they are not all equal, there exist consecutive distinct thresholds $g_1 < g_2$. For any $G \in [g_1, g_2)$, at least one leg with threshold g_1 binds and at least one leg with threshold at least g_2 does not bind. The binding set is therefore strict and nonempty on an interval with positive length. Equal crossing levels require equality of the ratios $(B_{j,s}(A) - c_{j,s}(A))/v_j$, which is destroyed by generic perturbations of loadings, costs, or caps. \square

Example 1 (Two-option Vega partial binding). For this example, fix a quote side s whose fills increase exposure in the direction of the local Greek coordinate below. Thus the displayed binding sets are side-specific leg sets $\mathcal{M}_s(q, A) \subseteq \{1, 2\}$; the corresponding side-level set is $\{(j, s) : j \in \mathcal{M}_s(q, A)\} \subseteq \mathcal{K}$. Consider two options with one shared volatility factor and local Vega loadings

$$v = (v_1, v_2), \quad v_1 > v_2 > 0.$$

The loadings are local quantities at the current state, not global constants. Let aggregate Vega exposure be

$$G(q) = v_1 q_1 + v_2 q_2.$$

Suppose the relevant side pressures are represented by the first-order pressure slice

$$p_j(q) = \alpha v_j G(q),$$

with $\alpha > 0$. This is the local form generated, for example, by a quadratic continuation-value approximation in the shared Vega coordinate. Assume common adverse-selection cost $c(A)$ and common cap-binding threshold $B(A)$. Option j binds when

$$\alpha v_j G(q) + c(A) \geq B(A).$$

For instance, if the local continuation penalty is proportional to $G(q)^2$, then the fill-induced value difference is locally proportional to $v_j G(q)$. Define

$$G_j^*(A) = \frac{B(A) - c(A)}{\alpha v_j}.$$

If $B(A) > c(A)$, then

$$G_1^*(A) < G_2^*(A).$$

Therefore the binding set evolves as

$$\begin{aligned} G(q) < G_1^*(A) &\implies \mathcal{M}_s(q, A) = \emptyset, \\ G_1^*(A) \leq G(q) < G_2^*(A) &\implies \mathcal{M}_s(q, A) = \{1\}, \\ G(q) \geq G_2^*(A) &\implies \mathcal{M}_s(q, A) = \{1, 2\}. \end{aligned}$$

This example uses common $B(A)$, common caps, and common adverse-selection costs only to isolate the loading effect. With heterogeneous caps and costs, the same threshold formula produces richer and generally non-nested crossing orders. The example is not a full solution of the HJB; it is a reduced pressure representation that isolates the binding-set geometry.

Example 2 (Mixed-sign loadings and non-nested binding). Now take three legs on a fixed side s with loadings

$$v_1 = 2, \quad v_2 = 1, \quad v_3 = -3,$$

and pressure slice

$$P_{j,s}(G, A) = \alpha v_j G + c(A).$$

Assume common threshold $B(A) > c(A)$ and $\alpha = 1$. If $G > 0$, legs with positive v_j have positive pressure and leg 3 has negative pressure, so leg 3 is farther from binding despite having the largest loading magnitude. The thresholds for legs 1 and 2 are

$$G_{1,s}^*(A) = \frac{B(A) - c(A)}{2}, \quad G_{2,s}^*(A) = B(A) - c(A),$$

while leg 3 cannot bind for positive G . For $G < 0$, the sign reverses: leg 3 has positive

pressure and the positive-loading legs have negative pressure. Hence the binding order is not obtained by sorting $|v_j|$. It depends on the sign of the aggregate exposure, the sign of the leg loading, and the side-specific thresholds.

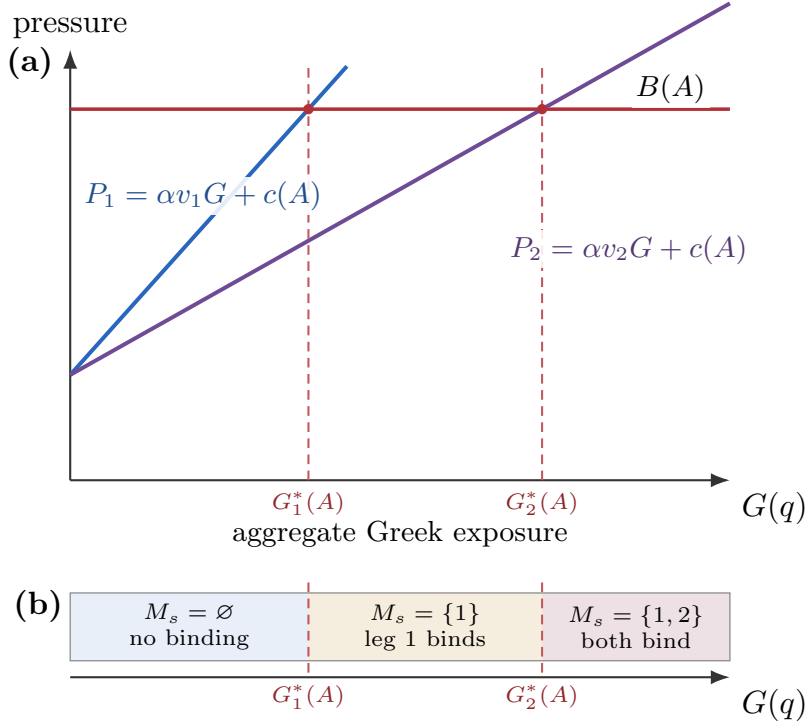


Figure 1: Partial cap binding from shared Vega exposure. Panel (a) shows the pressure crossings: the high-Vega leg reaches $B(A)$ at $G_1^*(A)$, while the lower-Vega leg reaches $B(A)$ at $G_2^*(A)$. Panel (b) shows the induced side-specific leg binding set $\mathcal{M}_s(q, A)$. The interval $[G_1^*(A), G_2^*(A))$ is the partial-binding region.

9. A Fixed-Point View of Constrained Implementation

The preceding results condition on a pressure vector at a single state. In the full dynamic problem, the binding set can feed back into the continuation value: capped or withdrawn sides alter future inventory distributions, which alter the value function, which alters pressures.

This distinction matters for notation. Statewise, the binding set is a subset

$$\mathcal{M}(z) \subseteq \mathcal{K}.$$

Dynamically, an implementation is a binding-set policy

$$\mathbf{m} : z \mapsto \mathbf{m}(z) \subseteq \mathcal{K}.$$

A compact way to express the constrained implementation problem is therefore as a fixed point in the space of such policies.

Let $V^{\mathbf{m}}$ denote the value function generated by an implementation rule that, at each state z , treats sides in $\mathbf{m}(z)$ as capped or subject to a withdrawal comparison and treats sides outside $\mathbf{m}(z)$ as interior. Let T map a candidate binding-set policy into the pressure field implied by that value function,

$$T(\mathbf{m}) = P[V^{\mathbf{m}}],$$

where $T(\mathbf{m})(z) \in \mathbb{R}^{|\mathcal{K}|}$ is a pressure vector at state z . Let Φ map any pressure field $P(\cdot)$ into the statewise threshold policy

$$\Phi(P)(z) = \{k \in \mathcal{K} : P_k(z) \geq B_k(A(z))\}.$$

A dynamically consistent constrained policy must satisfy

$$\mathbf{m} = \Phi(T(\mathbf{m})).$$

This formulation separates the two layers: T is the high-dimensional dynamic valuation map, while Φ is the sidewise threshold map studied above.

The fixed-point formulation also clarifies where deeper theory would enter. The natural domain is the lattice of maps from states into subsets of \mathcal{K} , ordered pointwise by inclusion. If T is monotone on this policy lattice, then $\Phi \circ T$ is monotone and Tarski-type arguments give existence of extremal fixed points. If T is a contraction under a metric on pressure fields or binding indicators, then the fixed point is unique. Establishing such conditions requires assumptions on the full HJB and is beyond the local binding-set stability result above, but the equation identifies the object that a complete dynamic constrained theory must characterize.

10. Cap Binding, Projection, and Withdrawal

The results above characterize cap binding. They become withdrawal results only under an additional no-fill or no-quote structure. If

$$\lambda_k(\delta_{k,\max}; A) > 0,$$

then the market maker is still quoting at the boundary and can still receive fills with positive probability. This is boundary quoting, not literal withdrawal.

To model withdrawal explicitly, add a no-quote payoff:

$$\max \left\{ \sup_{0 \leq \delta \leq \delta_{\max}} \lambda(\delta; A)(\delta - P), 0 \right\}.$$

The zero is a normalization of the incremental execution payoff. In the full HJB, the no-quote action removes the execution term while retaining the continuation value. The withdrawal boundary is the set of states in which the no-quote payoff dominates the best quoted-trade payoff. This boundary is conceptually distinct from the cap-binding boundary.

In applications, both may matter. A market maker may first move from an interior quote to a capped quote, then from a capped quote to a no-quote decision. The first transition is a projection or cap-binding event. The second is withdrawal. The present paper focuses on the first transition because it is the minimal constrained-implementation layer missing from an unconstrained interior approximation.

11. Conclusion

Quadratic approximations make high-dimensional option market-making problems computationally usable, but constrained implementations require more than a smooth quote map. They require a sidewise diagnostic that says which parts of the book remain interior and which parts have reached caps or no-quote comparisons.

This paper studies that diagnostic as a binding set. The scalar inverse-hazard cap condition is a tool. Applied side by side to the HJB pressure vector, it induces a state-dependent subset of option sides. The subset has monotone up-set structure, can expand through toxicity-driven

cascades, and can be partial or non-nested when Greek loadings are heterogeneous or mixed in sign.

The main approximation result is a stability theorem: the approximate binding set induced by a quadratic pressure field can disagree with the exact binding set only near exact threshold surfaces, within a tube whose width is governed by the pressure approximation error. Away from those margins, the approximate constrained-implementation diagnostic is exact.

For implementation, this suggests a clean division of labor. A quadratic engine supplies an interior pressure or quote map; a projection layer computes the binding set and records threshold margins; a no-quote layer, when present, performs a separate withdrawal comparison. The margin to the nearest threshold is then a natural stress diagnostic for option-book quoting systems.

Appendix A. Exact Global Threshold

For completeness, this appendix records the exact cap-binding threshold without the monotone-hazard condition. For a single side, the cap is globally optimal if and only if, for every $\delta < \delta_{\max}$,

$$\lambda(\delta_{\max}; A)(\delta_{\max} - P) \geq \lambda(\delta; A)(\delta - P).$$

Rearranging gives

$$P \geq \frac{\lambda(\delta; A)\delta - \lambda(\delta_{\max}; A)\delta_{\max}}{\lambda(\delta; A) - \lambda(\delta_{\max}; A)}$$

whenever the denominator is nonzero. Therefore the exact threshold is

$$\bar{P}(A) = \sup_{0 \leq \delta < \delta_{\max}} \frac{\lambda(\delta; A)\delta - \lambda(\delta_{\max}; A)\delta_{\max}}{\lambda(\delta; A) - \lambda(\delta_{\max}; A)}.$$

Because $\lambda(\cdot; A)$ is strictly decreasing, the denominator is positive for every $\delta < \delta_{\max}$. Under the monotone-hazard condition used in the main text, this supremum collapses to the endpoint hazard expression $B(A) = \delta_{\max} - 1/h(\delta_{\max}; A)$.

The cap is globally optimal if and only if

$$P \geq \bar{P}(A).$$

This result is robust but less transparent than the monotone-hazard expression. It is useful for checking the main theorem's assumptions and for extensions in which the fill curve is not log-concave.

References

- Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008. doi: 10.1080/14697680701381228.
- Bastien Baldacci, Philippe Bergault, and Olivier Guéant. Algorithmic market making for options, 2020a. arXiv:1907.12433.
- Bastien Baldacci, Joffrey Derchu, and Iuliia Manziuk. An approximate solution for options market-making in high dimension, 2020b. arXiv:2009.00907.
- Alexander Barzykin, Philippe Bergault, Olivier Guéant, and Malo Lemmel. Optimal quoting under adverse selection and price reading, 2025. arXiv:2508.20225.
- Philippe Bergault, David Evangelista, Olivier Guéant, and Douglas Vieira. Closed-form approximations in multi-asset market making. *Applied Mathematical Finance*, 28(2): 101–142, 2021. doi: 10.1080/1350486X.2021.1949359.
- Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1):415–444, 2014. doi: 10.1137/130911196.
- Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and High-Frequency Trading*. Cambridge University Press, Cambridge, 2015. ISBN 9781107091146.
- Sofiene El Aoud and Frédéric Abergel. A stochastic control approach for option market making, 2015. SSRN working paper.
- Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1): 71–100, 1985. doi: 10.1016/0304-405X(85)90044-3.
- Olivier Guéant. Optimal market making, 2017. arXiv:1605.01862.
- Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez-Tapia. Dealing with the inventory risk: A solution to the market making problem. *Mathematics and Financial Economics*, 7:477–507, 2013. doi: 10.1007/s11579-012-0087-0.

Fabien Guilbaud and Huyên Pham. Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, 13(1):79–94, 2013. doi: 10.1080/14697688.2012.708779.

Thomas Ho and Hans R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, 1981. doi: 10.1016/0304-405X(81)90020-9.

Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985. doi: 10.2307/1913210.

Vladimir Lucic and Alex Tse. Option market-making and vol arbitrage. *Risk*, 2024. Cutting Edge.